
W P S R L L (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 00:59:57 1999; MasPar time 1818.19 Seconds
Tabular output not generated. 1524.539 Million cell updates/sec

Title: >US-08-287-669-18
Description: (901-1900) from US08287669.seq (2 of 10)
Perfect Score: 1000

N.A. Sequence: 901 GTCGTCCTTGGTATCCTCAA.....TCGAAATTTGCTTTTATGTT 1900
Comp: CAGCAGGAACCATAGAGTT.....ACTTTTAAACAAAATACAA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158

Database: 1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_ov 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_v1

Database: 17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 11.164; Variance 7.504; scale 1.488

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description	Pred. No.
1	1000	100.0	7653	21	CELCE3A	Caenorhabditis elegans 0.00e+00
C 2	931	93.1	39908	21	CEC48D1	Caenorhabditis elegans 0.00e+00
C 3	925	92.5	314495	19	CEY67H2	Caenorhabditis elegans 0.00e+00
C 4	168	16.8	39908	21	CEC48D1	Caenorhabditis elegans 1.50e-78
C 5	163	16.3	207370	19	AC006798	Caenorhabditis elegans 1.59e-75
C 6	162	16.2	314495	19	CEY67H2	Caenorhabditis elegans 6.41e-75
C 7	160	16.0	306131	19	AC006874	Caenorhabditis elegans 1.04e-73
C 8	156	15.6	207370	19	AC006798	Caenorhabditis elegans 2.68e-71
C 9	156	15.6	299202	19	AC006792	Caenorhabditis elegans 2.68e-71
C 10	154	15.4	306131	19	AC006874	Caenorhabditis elegans 4.30e-70
C 11	148	14.8	32209	19	AC006663	Caenorhabditis elegans 1.75e-66
C 12	146	14.6	140702	20	AC006888	Caenorhabditis elegans 2.77e-65
C 13	146	14.6	317781	20	AC006906	Caenorhabditis elegans 2.77e-65

C 14	145	14.5	7653	21	CELCE3A	Caenorhabditis elegans 1.10e-64
C 15	144	14.4	125590	21	CEY49E10	Caenorhabditis elegans 4.38e-64
C 16	139	13.9	16938	22	CELT25D3	Caenorhabditis elegans 4.30e-61
C 17	139	13.9	125590	21	CEY49E10	Caenorhabditis elegans 4.30e-61
C 18	139	13.9	274498	19	AC006765	Caenorhabditis elegans 1.70e-60
C 19	138	13.8	143092	21	CEY17G7B	Caenorhabditis elegans 1.70e-60
C 20	138	13.8	207139	19	CEY17G7B	Caenorhabditis elegans 1.70e-60
C 21	137	13.7	39752	22	CELC04F5	Caenorhabditis elegans 6.73e-60
C 22	137	13.7	206217	19	AC006754	Caenorhabditis elegans 6.73e-60
C 23	136	13.6	42724	21	CEY53A2	Caenorhabditis elegans 6.73e-60
C 24	134	13.4	297866	19	AC006794	Caenorhabditis elegans 4.14e-58
C 25	134	13.4	299202	19	AC006792	Caenorhabditis elegans 4.14e-58
C 26	133	13.3	262336	19	AC006786	Caenorhabditis elegans 1.63e-57
C 27	132	13.2	140702	20	AC006888	Caenorhabditis elegans 6.43e-57
C 28	132	13.2	143092	21	CEY17G7B	Caenorhabditis elegans 6.43e-57
C 29	132	13.2	207139	19	CEY17G7B	Caenorhabditis elegans 6.43e-57
C 30	131	13.1	267118	20	AC006889	Caenorhabditis elegans 2.53e-56
C 31	129	12.9	254877	20	AC006904	Caenorhabditis elegans 3.90e-55
C 32	129	12.9	292390	19	AC006900	Caenorhabditis elegans 3.90e-55
C 33	128	12.8	32209	19	AC006663	Caenorhabditis elegans 1.53e-54
C 34	126	12.6	297872	19	AC006738	Caenorhabditis elegans 2.35e-53
C 35	125	12.5	19952	22	CELY38C9A	Caenorhabditis elegans 9.19e-53
C 36	125	12.5	38643	21	CELT03F1	Caenorhabditis elegans 9.19e-53
C 37	125	12.5	40662	21	CEY32B12B	Caenorhabditis elegans 9.19e-53
C 38	125	12.5	191857	19	CEY32B12	Caenorhabditis elegans 9.19e-53
C 39	125	12.5	297872	19	AC006844	Caenorhabditis elegans 9.19e-53
C 40	124	12.4	40662	21	CEY32B12B	Caenorhabditis elegans 3.59e-52
C 41	124	12.4	191857	19	CEY32B12	Caenorhabditis elegans 3.59e-52
C 42	123	12.3	35776	21	CELY02G9	Caenorhabditis elegans 1.40e-51
C 43	122	12.2	95968	21	CEY47D3B	Caenorhabditis elegans 5.47e-51
C 44	122	12.2	197735	19	CEY66A7	Caenorhabditis elegans 5.47e-51
C 45	122	12.2	298195	19	AC006890	Caenorhabditis elegans 5.47e-51

ALIGNMENTS

1	CELCE3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS	Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
DEFINITION					
ACCESSION	L29052				
NID	9456416				
VERSION	L29052.1 GI:456416				
KEYWORDS	cell death protein; interleukin-1 beta converting enzyme.				
SOURCE	Caenorhabditis elegans (strain N2) DNA.				
ORGANISM	Caenorhabditis elegans				
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.				
AUTHORS	Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.				
TITLE	The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1b-converting enzyme				
JOURNAL	Cell 75, 641-652 (1993)				
MEDLINE	94061982				
FEATURES	Location/Qualifiers				
Source	1..7653				
	/organism="Caenorhabditis elegans"				
	/strain="N2"				
exon	/db_xref="taxon:6239"				
	2167..2366				
	/gene="ced-3"				
gene	/number=1				
	join(2322..2366,2430..2575,2854..3107,4303..4634,				
	5547..5760,5815..5942,6298..6537,7013..7075)				
	/gene="ced-3"				
CDS	join(2322..2366,2430..2575,2854..3107,4303..4634,				
	5547..5760,5815..5942,6298..6537,7013..7075)				
	/gene="ced-3"				
	/note="codes for a protein similar to mammalian interleukin-1b-converting enzyme"				
	/codon_start=1				
	/product="cell death protein"				

/protein_id="AA27982.1"		Db		1081		TTTTTTTCCGCGAAATTTGCAATAAACCGCCAAAAACCTTTCTCCAAATGTTTACGCAA		1140	
/db_xref="PIR:g456417"		Qy		1081		TTTTTTTCCGCGAAATTTGCAATAAACCGCCAAAAACCTTTCTCCAAATGTTTACGCAA		1140	
/db_xref="GI:456417"		Db		1141		TATATACAATCCATAGATAATCTTCTCAATGTTTATGATTTCTTCGACGACTTTCTCT		1200	
/translation="MMQRDRSLRLNIMFSSHLKVDLEVLIAKOVLSNDGMI		Qy		1141		TATATACAATCCATAGATAATCTTCTCAATGTTTATGATTTCTTCGACGACTTTCTCT		1200	
NSCGTVREKREIVKAVRRGDVAFDAFALSTGHEGLAEVLEPLARSVDSNAVEF		Db		1201		TCGTGTCTAACATCTTATTTTATAATATTTCCCGCTAAAAATTCGGATTTTGGAGTATTA		1260	
ECMSPASHRRSRLSPAGYTSPTVRHDSVSVSFTSIQDIYSRARSRSRALHS		Qy		1201		TCGTGTCTAACATCTTATTTTATAATATTTCCCGCTAAAAATTCGGATTTTGGAGTATTA		1260	
SDRHNSPPVNAFPSPSSANSFTGCCSLGYSSRRNRKFSKASGPTQYIFHEEDMN		Db		1261		ATTATTCGTAATAATATATATAGCAGCGAAACCTACTTAAATAATGTTAAAGCTCCCTT		1320	
FVDAPTLRSVDEKTMVNFSSPRGMLIINNEHFQEMPTNRNGTKADKNLNLFCM		Qy		1261		ATTATTCGTAATAATATATATAGCAGCGAAACCTACTTAAATAATGTTAAAGCTCCCTT		1320	
GYTVICKDLTGRGMLTIIRDFAKHSHGDSAILVILSHGEENVILGVDDIPISTHEI		Db		1321		TAAATCGGCTCGACATTCGTTATTAAGGAATCACAAAATTCGAGAAATCGCTACGCGC		1380	
YDLLANAPRLANKPKIVFOACRGERDNGFPVLDSDVGVPAFLRGWDNRDGPLF		Qy		1321		TAAATCGGCTCGACATTCGTTATTAAGGAATCACAAAATTCGAGAAATCGCTACGCGC		1380	
NLCGVRPQVQWRKPKPSADLILRYATTAAQVSWNSARGSWFIQAVCEVFSTHAK		Db		1381		AACATATTTGACGGCAAAATATCTCGTAGCGAAACCTACAGTAATCTTTAAATGACTAC		1440	
DMVVELLTVENKVKVACGFOTSGSNNILKOMPETSKLLKFFWFPEARNSAV"		Qy		1381		AACATATTTGACGGCAAAATATCTCGTAGCGAAACCTACAGTAATCTTTAAATGACTAC		1440	
/gene="ced-3"		Db		1441		TGTAGCGCTGTGCTCGATTTACGGGCTCAATTTTGAATAATAATTTTTTTTCGAATTT		1500	
/number=1		Qy		1441		TGTAGCGCTGTGCTCGATTTACGGGCTCAATTTTGAATAATAATTTTTTTTCGAATTT		1500	
/number=2		Db		1501		TGATAACCCGTAATCGTCACACGCTACAGTAGTCATTTAAAGGATTAAGTACTAGTCTA		1560	
/gene="ced-3"		Qy		1501		TGATAACCCGTAATCGTCACACGCTACAGTAGTCATTTAAAGGATTAAGTACTAGTCTA		1560	
/number=3		Db		1561		GCTACGAGATATTTTTCGCGCCCAATATGACTGTAATACGCAATTCCTGAAATTTTGTT		1620	
/gene="ced-3"		Qy		1561		GCTACGAGATATTTTTCGCGCCCAATATGACTGTAATACGCAATTCCTGAAATTTTGTT		1620	
/number=4		Db		1621		TCCGTAATAATTTTCAAGAATTTTGGCATTTCCACTTTAAAGGCGCACAGATTTATCCA		1680	
/gene="ced-3"		Qy		1621		TCCGTAATAATTTTCAAGAATTTTGGCATTTCCACTTTAAAGGCGCACAGATTTATCCA		1680	
/number=5		Db		1681		ATGGGCTCTCGGCGCAAAAGTTTGTATGACATTTTAAATTCCTTCGATTTTAAATTC		1740	
/gene="ced-3"		Qy		1681		ATGGGCTCTCGGCGCAAAAGTTTGTATGACATTTTAAATTCCTTCGATTTTAAATTC		1740	
/number=6		Db		1741		AATTACTAAATTTTTCGTGAATTTTCTGTGTAATTTTAAATTCAGTTTCTTAATATT		1800	
/gene="ced-3"		Qy		1741		AATTACTAAATTTTTCGTGAATTTTCTGTGTAATTTTAAATTCAGTTTCTTAATATT		1800	
/number=7		Db		1801		TTCAGGCTGACAAACAGAAACAAACAAACAAACAAACAAACAAACAAATTTTAAATTC		1860	
/gene="ced-3"		Qy		1801		TTCAGGCTGACAAACAGAAACAAACAAACAAACAAACAAACAAATTTTAAATTCAGTTT		1860	
/number=8		Db		1861		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1861		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=9		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=10		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=11		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=12		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=13		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=14		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=15		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=16		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=17		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=18		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=19		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=20		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=21		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=22		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=23		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=24		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=25		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=26		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=27		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=28		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=29		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=30		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=31		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=32		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=33		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=34		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=35		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=36		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=37		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=38		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=39		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=40		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=41		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=42		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=43		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=44		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=45		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=46		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=47		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=48		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=49		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=50		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=51		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=52		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=53		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=54		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=55		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=56		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=57		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=58		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=59		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=60		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=61		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=62		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=63		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=64		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=65		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=66		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=67		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=68		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/gene="ced-3"		Qy		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	
/number=69		Db		1900		TAAATAACGATTTTCTCATTTGAAAATTTGTTTATGTT		1900	

RESULT 2 CEC48D1 39908 bp DNA INV 23-NOV-1998
 LOCUS Caenorhabditis elegans cosmid C48D1, complete sequence.
 DEFINITION
 ACCESSION 281049
 NID 91627677
 VERSION 281049.1 GI:1627677
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 39908)
 AUTHORS Burton, J.
 DIRECT SUBMISSION
 TITLE Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
 JOURNAL jes@sanger.ac.uk or rw@nemastode.wustl.edu

Cp 1660 TTAAAGTGGATGCAAAATCTTTGAAATTTATTCGGAACACAAATTCAGAAATG 1601
 Db 11754 CGTATTACGATCATATTTTGGCGGCAAAATATCTCTAGCTAGAACTACAGTAATCCTT 11813
 Cp 1600 CGTATTACG-TCATATTTGGCGGCAAAATATCTCTAGCTAGAACTACAGTAATCCTT 1542
 Db 11814 TAAATGACTACCGTAGCGCTTTGTGACGATTTACGGGTGTGCAAAATTCGAAAAAAT 11873
 Cp 1541 TAAATGACTACTGTAGG-TGTGACGATTTACGGGTGTGCAAAATTCGAAAAAAT 1483
 Db 11874 TATTTTCAAAATTTGACCGCTTAAATCGACACAAAGCGCTACAGTAGCTATTTAAAGAAAT 11933
 Cp 1482 TATTTTCAAAATTTGACCGCTTAAATCGACACAAAGCGCTACAGTAGCTATTTAAAGAAAT 1423
 Db 11934 ACTGTAGTTTTCGCTACGATATTTTGGCGCTCAAAATATGTTGGCAGTAGCGATTCCTC 11993
 Cp 1422 ACTGTAGTTTTCGCTACGATATTTTGGC-CGTCAAAATATGTTGGCAGTAGCGATTCCTC 1364
 Db 11994 AGAATTTTGTGATTCCTTAAATGAGTAATGTCGAGCGGATTTAAAG-A-CTTT-ACCAT 12050
 Cp 1363 AGAATTTTGTGATTCCTTAAATGAGTAATGTCGAGCGGATTTAAAGAGCTTTTACCAT 1304
 Db 12051 TTTTGTAGTTTTCGCTGCTATTTATGATATTTTACGATAAATTAATACTCAAAATCGG 12110
 Cp 1303 TTTTGTAGTTTTCGCTGCTATTTATGATATTTTACGATAAATTAATACTCAAAATCGG 1244
 Db 12111 AATTTAGCGGAATATTATAAAATTAAGATTTAGCAGCAAGAGAAAGTGTGCGAA 12170
 Cp 1243 AATTTAGCGGAATATTATAAAATTAAGATTTAGCAGCAAGAGAAAGTGTGCGAA 1184
 Db 12171 GAAATCAAAATTCGAGAGATTTCTTTATGATTTGATATATTTGCGTAACAAATTTGGA 12230
 Cp 1183 GAAATCAAAATTCGAGAGATTTCTTTATGATTTGATATATTTGCGTAACAAATTTGGA 1124
 Db 12231 GAAAGTTTTCGCGGTTTATTTGCAAAATTCGCGGAAAGAAAGTTTATGATGGCTCGT 12290
 Cp 1123 GAAAGTTTTCGCGGTTTATTTGCAAAATTCGCGGAAAGAAAGTTTATGATGGCTCGT 1064
 Db 12291 AAGACTGCGGATGAGAAAGAAAGTGGATTTGAGCAGCAAGCAATATGATCACCCTCTG 12350
 Cp 1063 AAGACTGCGGATGAGAAAGAAAGTGGATTTGAGCAGCAAGCAATATGATCACCCTCTG 1004
 Db 12351 GACAGTTTCAATTTCTAAACGATAATTTGAGCGGAG-CAGGTGCGATCAGGAGAGTGT 12409
 Cp 1003 GACAGTTTCAATTTCTAAACGATAATTTGAGCGGAGAGTGTGCGATCAGGAGAGTGT 944
 Db 12410 ACCGAAACAAACCGGACAGTTTGAGGATACCAAGGACAC 12452
 Cp 943 ACCGAAACAAACCGGACAGTTTGAGGATACCAAGGACAC 901

3
 CEX67H2 314495 bp DNA HTG 04-MAR-1999
 Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
 clone Y67H2, WORKING DRAFT SEQUENCE.
 AL022475
 ACCESSION
 NID 94469034
 VERSION AL022475.3 GI:4469034
 KEYWORDS HTG; HTGS; PHASE1.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 McMuray, A.
 Direct Submission
 Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,
 Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
 Washington University, St. Louis, MO 63110, USA. E-mail:
 jes@sanger.ac.uk or wren@nematoe.wustl.edu
 On Mar 22, 1999 this sequence version replaced gi:4468145.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source

Location/Qualifiers
 1. .314495
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="Y67H2"

BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others
 ORIGIN

Query Match 92.5%; Score 925; DB 19; Length 314495;
 Best Local Similarity 98.3%; Pred. No. 0.00e+00;
 Matches 987; Conservative 0; Mismatches 8; Indels 9; Gaps 9;

Db 241112 AACATAAACAACAATTTCAATGAGAAATCCTTATTTTAAATTTGAAACTGATTTTAA 241171
 Cp 1900 AACATAAACAACAATTTCAATGAGAAATCCTTATTTTAAATTTGAAACTGATTTTAA 1841
 Db 241172 AATGTTTGTGTTTGTGTTTCTGTTGTCAGCTCGAAATATTAGAAACTGATTTT 241231
 Cp 1840 AATGTTTGTGTTTGTGTTTCTGTTGTCAGCTCGAAATATTAGAAACTGATTTT 1781
 Db 241232 AAAAATTTTACAGAAAAATTCACGAAAAATTTTAGTAATTTGAATTTAAATGCAAGGAGA 241291
 Cp 1780 AAAAATTTTACAGAAAAATTCACGAAAAATTTTAGTAATTTGAATTTAAATGCAAGGAGA 1721
 Db 241292 ATTTAAAGTCTATCAACTTTTTCGTCGCGAGACCAATTTGAATAATTC-GTGCGCC 241350
 Cp 1720 ATTTAAAGTCTATCAACTTTTTCGTCGCGAGACCAATTTGAATAATTC-GTGCGCC 1661
 Db 241351 TTAAAGAGGAATGCCAAATCTTTGTAATTTATTCGGAACACAAAAATTCAGAGAATG 241410
 Cp 1660 TTAAAGTGGAAATGCCAAATCTTTGTAATTTATTCGGAACACAAAAATTCAGAGAATG 1601
 Db 241411 CGTATTACGATCATATTTTGGCGGCAAAATATCTCTAGCTAGAACTACAGTACATCT 241470
 Cp 1600 CGTATTACG-TCATATTTTGGCGGCAAAATATCTCTAGCTAGAACTACAGTA-ATCCT 1543
 Db 241471 TTAATGACTACCGTACGCTTTGTGACGATTTACGGGTGTGCAAAATTCGAAAAAATA 241530
 Cp 1542 TTAATGACTACTGTACG-TTGTGACGATTTACGGGTGTGCAAAATTCGAAAAAATA 1484
 Db 241531 TTATTTTCAAAATTTGAGCCGTAATTCGACACAAAGCGCTACAGTAGTTCATTTAAAGAAAT 241590
 Cp 1483 TTATTTTCAAAATTTGAGCCGTAATTCGACACAAAGCGCTACAGTAGTTCATTTAAAGAAAT 1424
 Db 241591 TACTGTAGTTTTCGCTACGAGATATTTTGGCGGTCAAAATATGTTGGCGAGTACGATTCCT 241650
 Cp 1423 TACTGTAGTTTTCGCTACGAGATATTTTGGC-CGTCAAAATATGTTGGCGAGTACGATTCCT 1365
 Db 241651 CAGAAATTTGTGATTCCTTAAATACGATAATTCGAGCCGATTTAAAG-A-CTTT-ACCA 241707
 Cp 1364 CAGAAATTTGTGATTCCTTAAATACGATAATTCGAGCCGATTTAAAGAGAGCTTTTACCA 1305
 Db 241708 TTTTGTAGTTTTCGCTGCTATTTATGATAATTTTACGATAAATTAATACCAAAATTCG 241767
 Cp 1304 TTTTGTAGTTTTCGCTGCTATTTATGATAATTTTACGATAAATTAATACCAAAATTCG 1245
 Db 241768 GAATTTTACGGAATATTTATAAAATAAGATTTAGCAGCAAGAGAAAGTGTGCTGCGA 241827
 Cp 1244 GAATTTTACGGAATATTTATAAAATAAGATTTAGCAGCAAGAGAAAGTGTGCTGCGA 1185
 Db 241828 AGAAATCAAAACATTTGAGAAGATATTTTATGATTTATTTGCTAACAAATTTGG 241887
 Cp 1184 AGAAATCAAAACATTTGAGAAGATATTTTATGATTTGCTATATTTGCTAACAAATTTGG 1125

Db	241888	AGAAAGTTTGGCCGGTTTATTTGCAAAATTCGCGGAAAAAAGTTTATGATGGCTCG	241947	
Cp	1124	AGAAAGTTTGGCCGGTTTATTTGCAAAATTCGCGGAAAAAAGTTTATGATGGCTCG	1065	
Db	241948	TAAGCAATGCGGATGAGAAAAAGAGTGGATTTAGCAGCAGCAATATGAGTACCCATCT	242007	
Cp	1064	TAAGCAATGCGGATGAGAAAAAGAGTGGATTTAGCAGCAGCAATATGAGTACCCATCT	1005	
Db	242008	GGACAGTTACATTTCTAAACGATATTTAGACGGAG-CAGGTGGCATCACGGAAGAGTG	242066	
Cp	1004	GGACAGTTACATTTCTAAACGATATTTAGACGGAG-CAGGTGGCATCACGGAAGAGTG	945	
Db	242067	TACCGAAAAACACCGGACAAAGTTGAGGATACCAAGGACGAC	242110	
Cp	944	TACCGAAAAACACCGGACAAAGTTGAGGATACCAAGGACGAC	901	
RESULT	4	CBC48D1	39908 bp	DNA
LOCUS		Caenorhabditis elegans	cosmid C48D1	complete sequence.
DEFINITION		Caenorhabditis elegans		
ACCESSION		281049		
NID		g1627677		
VERSION		281049.1	GI:1627677	
KEYWORDS		HTG.		
SOURCE		Caenorhabditis elegans.		
ORGANISM		Caenorhabditis elegans		
REFERENCE		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Pelodidae; Caenorhabditis.		
AUTHORS		Burton, J.		
TITLE		Direct Submission		
JOURNAL		Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nemastode.wustl.edu		
REFERENCE		2 (bases 1 to 39908)		
AUTHORS		Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Lalister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Shownkeen, R., Smailson, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.		
TITLE		2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans		
JOURNAL		Nature 368 (6466), 32-38 (1994)		
MEDLINE		94150718		
COMMENT		Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=C48D1 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. This sequence is the entire insert of clone C48D1. The true right end of clone F58D2 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence 281093. The end of this sequence (33020..39908) overlaps with the start of sequence 282274. Location/Qualifiers 1..39908		

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(994..4439)
/gene="C48D1.1"
complement(join(994..1020,1345..1951,2283..2580,2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/gene="C48D1.1"
/note="similar to C.elegans zyg-11 like protein; cDNA EST EMBL:214909 comes from this gene; cDNA EST EMBL:101768 comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="PID:el344741"
/db_xref="PID:g3875048"
/db_xref="GI:3875048"
/db_xref="SPTREMBL:O02227"
/translation="MTKPLHDIACQKIAEFVQSGYDNRFEELDSKSSNIVSYLMNR DKPNLKT IHLNVT KVTLPQTLDTENTARQLKQFSLKELKIGNFKSFCIFRKE TRVDIVRILEAVLNEETREKLOVLTHTKQKLMKRTAKIGNMLSLTSLTMSDVHI NEDDYQOLFQNFSLRSLDISGTLNLIAGISLKOELVAFLOFSTATVTELTDLFN LKHLRVLDVSQTRFNLSRNIVEQFTECKMELPELREMDCTRTSLSGDILECLLQ LSPNIISVTEAETSOYDNRVLTNTAFYHKSQVQALVYCLLFAKFEFDSGCMKAIKRSI SEDNLNIPSCITIKVQNTFDSISVQNGKLKLIIFGMDYSNGLCSNGVTLVSE ALMSFVLYKRMISEAYKVNISYNAVKLVNKLKGLDFDKISKTAMFSLERVRS KALTRASSVLCCQIHRMHSDELKSLSEKTEKVNRLVDILEWSVGOKNFKLSEFFEI TIKMTKSSKACTNNIQAIGISISLSCKMEKVESNDITTKCLELTLNLSALNVDLSQE LYTSNVNDKIGGANAISTEK"
complement(6258..11120)
/gene="C48D1.2"
complement(join(6258..6349,6876..7065,7420..7547,7602..7815,8728..9059,10248..10501,10780..10925,10989..11120))
/gene="C48D1.2"
/note="similar to Caspase recruitment domain, ICE-like protease (caspase) p20 domain., ICE-like protease (caspase) p10 domain.; cDNA EST EMBL:D75533 comes from this gene; cDNA EST yk198f10.5 comes from this gene; cDNA EST yk491a5.5 comes from this gene"
/codon_start=1
/protein_id="CAB02848.1"
/db_xref="PID:el344743"
/db_xref="PID:g3875050"
/db_xref="GI:3875050"
/db_xref="SPTREMBL:O02229"
/translation="MEQRRSLLENIMFSSHLKVDLEVLIAKOVNDSNDGMIN SCGTVRKRRRIKAVQRGVADFDALRSTGHEGLAEVLEPLARSVDSNAVEFE CPMSPAHRRSRALSPAGTTSPTVRHRSVSVSVSFTSYQDIYKRSRSLRSLHSS DRHNTSPPVNAFPSPQSSANSSTGSSLSGYSRNRKSKAGSPTQIIFHEEDMNF VDAPTISRVDEKTYRNFSPRGMLIINNEHFEQMTNRGTAKDNLTNLFRCMG YTVICKDNLGTGRMLTTRDFAKHSGDSAILVLHSGEENVIGVDDIDIPITHEIY DLLNAANRLANKPKIVFQACGERDNGFPVLDSVDGVPFAFLRGRDNGDGLFN FLGCVPOQVQWRKPKSQADILLATAYTAQVSVWRNSARGSWFIQACEVFEVTHAKD MDVVELLEVKNKVAIDIPPAQKVLGAGSKLCKLKFTRDSLPN"
complement(16940..21616)
/gene="C48D1.3"
complement(join(16940..17126,18419..18597,18985..19442,19553..19879,19925..19961,20333..20481,20789..20987,21037..21152,21373..21616))
/gene="C48D1.3"
/note="Similarity to Salmonella sodium/proline symporter (SW:PUTP_SALTY); cDNA EST EMBL:C13840 comes from this gene; cDNA EST EMBL:C11562 comes from this gene"
/codon_start=1
/protein_id="CAB02847.1"
/db_xref="PID:el344742"
/db_xref="PID:g3875049"
/db_xref="GI:3875049"
/db_xref="SPTREMBL:O02228"
/translation="NADLIGIVAIIVFVVLIVGVIGWAGRSKSKSESEAGAATEE VMLAGNIGTIVGIFTMTATVGGVINGTAALYNGLLGCOAPGVVAISVGMGLL FAKKREEGITIMLDPFQFWNFLEIFGTFFNFKLGRFLQLQIIEILDFQHKYG


```

8
RESULT AC006798 207370 bp DNA HTG 25-FEB-1999
LOCUS Caenorhabditis elegans clone Y51F8, WORKING DRAFT SEQUENCE, 30
DEFINITION unordered pieces.
ACCESSION AC006798
NID 94309785
VERSION AC006798.2 GI:4309785
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE 1 (bases 1 to 207370)
JOURNAL Waterston, R.H.
REFERENCE The sequence of Caenorhabditis elegans clone
AUTHORS 2 (bases 1 to 207370)
TITLE Unpublished
JOURNAL Waterston, R.H.
REFERENCE Direct Submission
AUTHORS Submitted (23-FEB-1999) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
COMMENT On Mar 1, 1999 this sequence version replaced gi:4263140.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1986: contig of 1986 bp in length
1987
2000: gap of unknown length
2001
4562: contig of 2562 bp in length
4563
4577: gap of unknown length
4578
7877: contig of 3301 bp in length
7878
7891: gap of unknown length
7892
10172: contig of 2281 bp in length
10173
10186: gap of unknown length
10187
12395: contig of 2209 bp in length
12396
12409: gap of unknown length
12410
13306: contig of 2897 bp in length
13307
15320: gap of unknown length
15321
17774: contig of 2454 bp in length
17775
17788: gap of unknown length
17789
19333: contig of 2145 bp in length
19334
19947: gap of unknown length
19948
24445: contig of 4497 bp in length
24459
27765: contig of 3307 bp in length
27766
27779: gap of unknown length
27780
31869: contig of 4090 bp in length
31870
31883: gap of unknown length
31884
37830: contig of 5953 bp in length
37837
37851: gap of unknown length
42798
42811: gap of unknown length
42812
46266: contig of 3455 bp in length
46267
46280: gap of unknown length
46281
49619: contig of 3339 bp in length
49620
49633: gap of unknown length
49634
52707: contig of 3074 bp in length
52708
52721: gap of unknown length
52722
56266: contig of 3545 bp in length
56267
56280: gap of unknown length
56281
59691: contig of 3411 bp in length
59692
59705: gap of unknown length
59706
63841: contig of 4136 bp in length
63842
63855: gap of unknown length
70264: contig of 6409 bp in length
70278: gap of unknown length
75486: contig of 5208 bp in length
70279

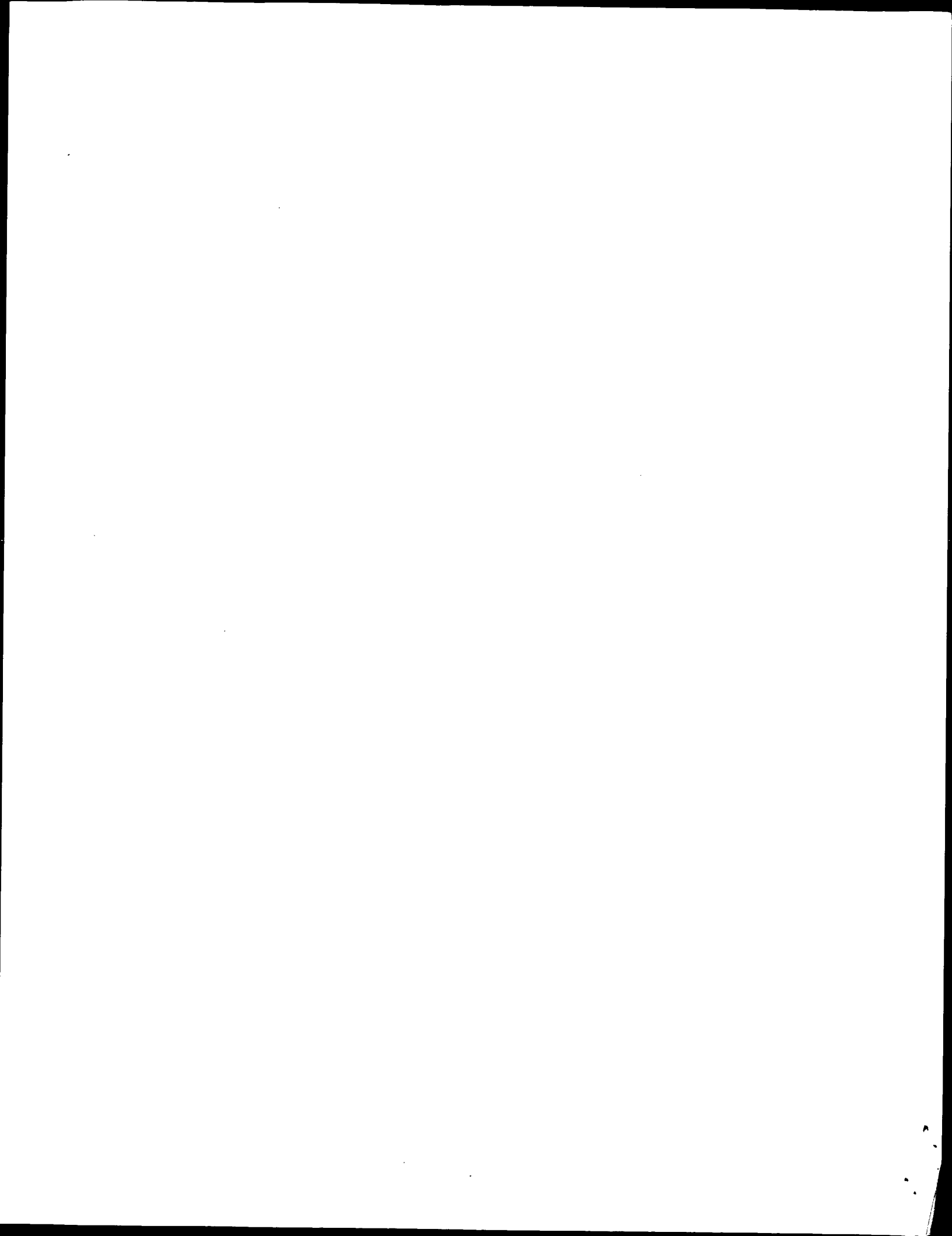
```

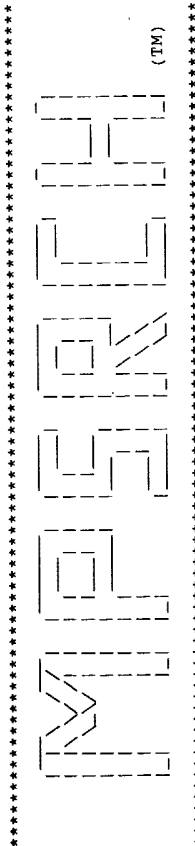
```

75487 75500: gap of unknown length
75501 80985: contig of 5485 bp in length
80986 80999: gap of unknown length
81000 85954: contig of 4955 bp in length
85955 85968: gap of unknown length
85969 92274: contig of 6306 bp in length
92275 92288: gap of unknown length
92289 98525: contig of 6237 bp in length
98526 98539: gap of unknown length
98540 103903: contig of 5364 bp in length
103904 103917: gap of unknown length
103918 110068: contig of 6151 bp in length
110069 110082: gap of unknown length
110083 122583: contig of 12501 bp in length
122584 122597: gap of unknown length
122598 147659: contig of 25062 bp in length
147660 147673: gap of unknown length
147674 207370: contig of 59697 bp in length.
FEATURES
    Location/Qualifiers
        1..207370
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /clone="Y51F8"
BASE COUNT 69298 a 36416 c 35867 g 65363 t 426 others
ORIGIN
Query Match 15.6%; Score 156; DB 19; Length 207370;
Best Local Similarity 82.2%; Pred. No. 2.68e-71;
Matches 244; Conservative 0; Mismatches 46; Indels 7; Gaps 5;
Db 79955 TATCAGCGGACACTAAATCTCTGAAATGGGTATTACACACCATATTTGACGCGCAAAAT 80014
QY 1342 TATTAAAGGAATCACAAAATCTGAGAAATGCTACTGCGCAACATATTTGACG-GCAAAAT 1400
Db 80015 ATCTCGTAGCGAAACTACAGTAATCTTTAAATGACTATTGTAGCGCTCTGTGTCGAA 80074
QY 1401 ATCTCGTAGCGAAACTACAGTAATCTTTAAATGACTACTGTAGCGCT--TGTGTCGAT 1458
Db 80075 TTCCGGGATCTGATTTTGTGAACTATTAATCATTTTCGGAAAAAGAGATCGTAAATC 80134
QY 1459 TTACGGG--CTCAATTTTGTGAAATATATTTTTCGAAATTTTTCGAAATCGTAAATC 1516
Db 80135 GACAAAGGGCTACAGTAGTCATTAAAGAGATTATTTGACTATTTTCGAAAAAGAGATCGTAAATC 80194
QY 1517 GTCAACAG-CTACAGTAGTCATTAAAGGATTACTGTAGTTCTAGCTACGAGATATTTT 1575
Db 80195 CGCGCTCAAAATATTTGTGCAAAACGCAATCTTTGAAATTTTGTTCCTCGTAAATAT 80251
QY 1576 CGCGCCCAATATGAC-TGTAATACGCATCTCTGAATTTTGTGTTTCGTAATAT 1631
RESULT 9
LOCUS AC006792 299202 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone Y50C1a, WORKING DRAFT SEQUENCE, 10
ACCESSION AC006792
NID 94263146
VERSION AC006792.1 GI:4263146
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE 1 (bases 1 to 299202)
JOURNAL Waterston, R.H.
REFERENCE The sequence of Caenorhabditis elegans clone
AUTHORS 2 (bases 1 to 299202)
TITLE Unpublished
JOURNAL Waterston, R.H.
REFERENCE Direct Submission
AUTHORS Submitted (23-FEB-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```


100





Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 03:00:19 1999; MasPar time 219.29 Seconds
Tabular output not generated.
977.118 Million cell updates/sec

Title: >US-08-287-669-18
Description: (901-1900) from US08287669.seq (2 of 10)
Perfect Score: 1000
N.A. Sequence: 901 GTGCTCCTTGCTATCCTCAA.....TGAAATTTGTTTATGTT 1900
Comp: CAGCAGGACCATAGGAGTT.....ACTTTTAAACACAAATACAA

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 271905 seqs, 107135622 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60
Statistics: Mean 8.878; Variance 6.553; scale 1.355

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	998	99.8	7653	24	T38196	0.00e+00
2	996	99.6	7653	9	Nematode Ced-3 gene.	0.00e+00
3	992	99.2	7653	9	ced-3 (G2487A) gene.	0.00e+00
4	992	99.2	7653	9	ced-3 (C5940T) gene.	0.00e+00
5	992	99.2	7653	9	ced-3 (G6372A) gene.	0.00e+00
6	992	99.2	7653	9	ced-3 (G6536A) gene.	0.00e+00
7	992	99.2	7653	9	ced-3 (G64739) gene.	0.00e+00
8	992	99.2	7653	9	ced-3 (C7020T) gene.	0.00e+00
9	992	99.2	7653	9	ced-3 (G6434T) gene.	0.00e+00

10	992	99.2	7653	9	Q64743	ced-3 (C6485T) gene.	0.00e+00
11	992	99.2	7653	9	Q64740	ced-3 (C6322T) gene.	0.00e+00
12	992	99.2	7653	9	Q54666	ced-3 gene.	0.00e+00
13	992	99.2	7653	9	Q64737	Nematode Ced-3 gene.	0.00e+00
14	145	14.5	7653	24	T38196	ced-3 (G5757A) gene.	1.27e-62
15	143	14.3	7653	9	Q64744	ced-3 (G6536A) gene.	1.64e-61
16	143	14.3	7653	9	Q64743	ced-3 (C6485T) gene.	1.64e-61
17	143	14.3	7653	9	Q64737	ced-3 (G5757A) gene.	1.64e-61
18	143	14.3	7653	9	Q64745	ced-3 (C7020T) gene.	1.64e-61
19	143	14.3	7653	9	Q64740	ced-3 (C6322T) gene.	1.64e-61
20	143	14.3	7653	9	Q64739	ced-3 (G6297A) gene.	1.64e-61
21	143	14.3	7653	9	Q64738	ced-3 (C5940T) gene.	1.64e-61
22	143	14.3	7653	9	Q64736	ced-3 (G2487A) gene.	1.64e-61
23	143	14.3	7653	9	Q64742	ced-3 gene.	1.64e-61
24	143	14.3	7653	9	Q54666	ced-3 (G6372A) gene.	1.64e-61
25	143	14.3	7653	9	Q64741	ced-3 gene.	1.64e-61
26	141	14.1	7653	9	Q54401	ced-3 gene.	2.11e-60
27	81	8.1	6560	9	Q54629	Genomic region contai	9.82e-28
28	76	7.6	6560	9	Q54629	Genomic region contai	4.29e-25
29	41	4.1	91	9	Q51746	Oligonucleotide probe	2.06e-07
30	40	4.0	204	1	N81164	Base substituted E.co	4.45e-05
31	36	3.6	204	1	N81164	Oligonucleotide probe	4.45e-05
32	36	3.6	204	1	N81164	Base substituted E.co	4.45e-05
33	35	3.5	91	46	V44650	Mammalian DNA replica	1.27e-04
34	33	3.3	91	46	V44650	Mammalian DNA replica	1.02e-03
35	32	3.2	501	3	N50023	Sequence encoding new	2.84e-03
36	30	3.0	114	12	Q70469	Generic DNA sequence	2.14e-02
37	30	3.0	501	3	N50026	Sequence encoding new	2.14e-02
38	30	3.0	501	3	N50024	Sequence encoding new	2.14e-02
39	30	3.0	501	3	N50027	Sequence encoding new	2.14e-02
40	28	2.8	114	12	Q70467	Generic DNA sequence	1.54e-01
41	28	2.8	219	19	T20616	Human gene signature	1.54e-01
42	28	2.8	501	3	N50031	Sequence encoding new	1.54e-01
43	28	2.8	501	3	N50025	Sequence encoding new	1.54e-01
44	28	2.8	501	3	N50033	Sequence encoding new	1.54e-01
45	28	2.8	501	3	N50029	Sequence encoding new	1.54e-01

ALIGNMENTS

RESULT 1	ID	T38196 standard; DNA; 7653 BP.
AC	T38196;	
DT	17-DEC-1996 (first entry)	
DE	Nematode Ced-3 gene.	
KW	Ced-3; interleukin-1 beta converting enzyme; ICE; protease;	
KW	cell death; apoptosis; neutral degeneration; inflammation;	
OS	antiinflammatory; ds.	
OS	Caenorhabditis elegans.	
PH	Key	Location/Qualifiers
FT	repeat_unit	1356..1472
FT		/*tag= a
FT	repeat_unit	1490..1614
FT		/*tag= b
FT	exon	2167..2366
FT		/*tag= c
FT	allele	/codon_start= 2232..2234
FT		2310
FT		/*tag= d
FT		/label= T(n1040)
FT	intron	/note= "causes L27F mutation"
FT		2367..2429
FT		/*tag= e
FT	exon	/label= Intron-1
FT		2430..2575
FT		/*tag= f
FT	allele	2487
FT		/*tag= g
FT		/label= A(n718)
FT	intron	/note= "causes G65R mutation"
FT		2576..2853

```

FT      /*tag= h
FT      /label= Intron-2
FT      2854...3107
FT      /*tag= i
FT      3108...4302
FT      /*tag= j
FT      /label= Intron-3
FT      3126...3243
FT      /*tag= k
FT      /label= Repeat-1
FT      3329...3396
FT      /*tag= l
FT      /label= Repeat-1
FT      3487...3759
FT      /*tag= m
FT      /label= Repeat-2
FT      3782...4070
FT      /*tag= n
FT      /label= Repeat-2
FT      4303...4633
FT      /*tag= o
FT      4634...5546
FT      /*tag= p
FT      /label= Intron-4
FT      4688...4719
FT      /*tag= q
FT      /label= Repeat-3
FT      5221...5330
FT      /*tag= r
FT      /label= Repeat-3
FT      5546...5760
FT      /*tag= s
FT      5757
FT      /*tag= t
FT      /label= A(n2433)
FT      /note= "causes G360S mutation"
FT      5761...5814
FT      /*tag= u
FT      /label= Intron-5
FT      5815...5942
FT      /*tag= v
FT      5940
FT      /*tag= w
FT      /label= T(n1165)
FT      /note= "creates premature stop codon at 403"
FT      5943...6297
FT      /*tag= x
FT      /label= Intron-6
FT      6062...6138
FT      /*tag= y
FT      /label= Repeat-4
FT      6298...6537
FT      /*tag= z
FT      6322
FT      /*tag= aa
FT      /label= T(n1949)
FT      /note= "creates premature stop codon at 412"
FT      6372
FT      /*tag= ab
FT      /label= A(n1286)
FT      6434
FT      /*tag= ac
FT      /label= T(n1129,n1164)
FT      /note= "causes A449V mutation"
FT      6485
FT      /*tag= ad
FT      /label= T(n2430)
FT      /note= "causes A466V mutation"
FT      6535
FT      /*tag= ae
FT      /label= A(n2426)
FT      /note= "causes E483K mutation"
FT      6538...7012

```

```

FT      /*tag= af
FT      /label= Intron-7
FT      6567...6625
FT      /*tag= ag
FT      /label= Repeat-5
FT      6905...6965
FT      /*tag= ah
FT      /label= Repeat-5
FT      7013...7653
FT      /*tag= ai
FT      /note= "in-frame stop codon at 7073-7075"
FT      7020
FT      /*tag= aj
FT      /label= T(n1163)
FT      /note= "causes S486F mutation"
FT      WO9625946-A1.
FT      29-AUG-1996.
FT      23-FEB-1996; US-394189.
FT      24-FEB-1995; US-394189.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Horvitz HR, Shaham S, Yuan J;
FT      WPI; 96-425082/42.
FT      P-PSDB; R98754.
FT      Ced-3 and human interleukin 1-beta convertase genes and proteins -
FT      useful to treat inflammation and diseases characterised by cell
FT      death
FT      Claim 18; Fig 3; 139pp; English.
FT      The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
FT      mapping RFUs and chromosome walking, and genomic DNA cloned
FT      in plasmid pJ107 was sequenced. EMS-induced alleles were also
FT      sequenced. The gene codes for a cell death protein (R98754) that
FT      is structurally similar to human interleukin-1 beta converting
FT      enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
FT      cysteine protease like ICE and that ICE may be a human equivalent
FT      of the nematode cell death gene. The ced-3 gene can be used as a
FT      probe or in the prodn. of Ced-3 protein and novel drugs for
FT      enhancing or inhibiting the activity of ICE, ced-3 and related
FT      genes for the treatment of inflammatory diseases and/or diseases
FT      caused by cell death. Novel inhibitors of ced-3 activity include
FT      portions of the ced-3 gene and its product.
FT      Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;
FT      99.8%; Score 998; DB 24; Length 7653;
FT      Best Local Similarity 99.9%; Pred. NO. 0.00e+00;
FT      Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db      901 gtcgtctctggtatctcaactgtcccggttttttcggtacactttccggtatgc 960
QY      |||||||
QY      901 GTCGTCTTGGTATCTCTCAACTGTCCCGGTTTGTTCGGTACACTCTCCCGTGATGC 960
Db      961 cactgtctcgtctcctcaattatctgttagaaatgtgaactgtccagatgggtgactcata 1020
QY      |||||||
QY      961 CACTGTCTCCGCTCAATTATCTGTTAGAAATGTGAAGTGTCCAGATGGGTGACTCATA 1020
Db      1021 ttgctgtctcaatccactttttttctcatcgccagctcttaagccatcataaac 1080
QY      |||||||
QY      1021 TTGCTGCTCTCAATCCACTTTCTTCTCATCGCGAGTCTTACGAGCCCATATAAC 1080
Db      1081 ttttttttcgcgcaatttgcaataaacgcgcaaaactttctccaattgttacgcaa 1140
QY      |||||||
QY      1081 TTTTTCGCGCAATTGCAATTAACGCGCAAAACTTCTCCAAATGTTACGAA 1140
Db      1141 tatataatccataagaatatcttctcaatgtttatgatatttcttcgcagcactttctct 1200
QY      |||||||
QY      1141 TATATACAATCCATAAGAATATCTTCTCAATGTTTATGATTTCTTCGCAGCACTTTCTCT 1200
Db      1201 tcgtgtctcaatcttatttttataatttcgcgtataaaattccgatttttgagtatta 1260
QY      |||||||
QY      1201 TCGTGTCTACATCTTATTTTATATATTTCGCTAAATATCCGATTTTGTAGTATTA 1260
Db      1261 attatcgtataaattatcataatagcaccgaaactactataaagtgttaaaagctcttt 1320
QY      |||||||
QY      1261 ATTTATCGTAAATATATCATAATAGCACCAGAAACTACTATAAATGTTAAAGCTCCTTT 1320

```

Db	1321	taaatcgctcgacattatctgtatttaaggaatcacaaaattcttgagaatcgctactgcgc	1380	FT	mutation	/number= Exon_2
Qy	1321	TAAATCGGCTCGACATTATCGTATTAGGAATCACAAAATCTGAGAATCGTACTGCOC	1380	FT	intron	/tag= g
Db	1381	aacatattgacggcaaaaatctctgtagcgaaaaatacagtaattctttaaataactact	1440	FT	exon	/note= "G>A, from allele n718"
Qy	1381	AACATATTGACGGCAAAAATCTCTGAGGAAATACAGTAATCTTTAAATGACTAC	1440	FT	exon	2576..2853
Db	1441	tgtagcgtgtgctgattacggctcaatttttgaaataatttttttcgaattt	1500	FT	intron	/tag= h
Qy	1441	TGTAGCGTGTGTCGATTTACGGCTCAATTTTGAATAATATTTTTTTTCGAATTT	1500	FT	intron	/number= Intron_2
Db	1501	tgataacccgttaactgcacacgctacagtagtgcattttaaggattactgtagtctta	1560	FT	repeat_unit	2854..3107
Qy	1501	TGATAACCCGTAATTCGTCACACGCTACAGTAGTCATTTTAAAGGATTACTGTAGTTCTA	1560	FT	repeat_unit	/tag= i
Db	1561	gctacgagatatttgcgcgcaaaatgactgtacgtacacattctctgaatttctgttt	1620	FT	repeat_unit	/number= Exon_3
Qy	1561	GCTACGAGATATTTGCGCGCAAAATGACTGTAAATACGATTCCTGGAATTTTGTTT	1620	FT	repeat_unit	3108..4302
Db	1621	tcctgaataatttcacaagattttggcattccactttaaggcgccagcagattattcca	1680	FT	repeat_unit	/tag= j
Qy	1621	TCCGTAAATTAATTTACAGATTTTGGCATTCACATTTTAAAGGCGCACAGATTATCCCA	1680	FT	repeat_unit	/number= Intron_3
Db	1681	atgggtctcgccagcaaaaagtttgatagacttttaaaattctctgcatttttaattc	1740	FT	repeat_unit	3126..3243
Qy	1681	ATGGGTCTCGCCAGCAAAAAGTTTGATAGACTTTTAAATTCCTTCGATTTTAAATTC	1740	FT	repeat_unit	/tag= k
Db	1741	aattactaaaatttgcgtgaatttttctgttaaaatttttaaaatcagttttctaatatt	1800	FT	repeat_unit	/rpt_type= INVERTED
Qy	1741	AATTACTAAAATTTTCGTGAATTTTTCGTGTAATTTTAAATCAGTTTCTTAATATT	1800	FT	repeat_unit	/note= "Inverted w.r.t. repeat at 3126-3243"
Db	1801	ttccaggctgacacacagcaaaaacacacacacacatttttaaaatcagttttccaaat	1860	FT	exon	/tag= l
Qy	1801	TTCCAGGCTGACACACAGCAAAAACACACACACAAATTTTAAATCAGTTTCCAAAT	1860	FT	exon	/number= Exon_4
Db	1861	taaaaataacgatttctctgaaattgtgttttatgtt	1900	FT	mutation	/tag= p
Qy	1861	TAAAAATAACGATTCTCTATTGAAAATGTGTTTATGTT	1900	FT	mutation	/number= Intron_4
RESULT	2			FT	repeat_unit	4688..4719
ID	Q54401	standard; DNA; 7653 BP.		FT	repeat_unit	/tag= q
AC	Q54401;			FT	repeat_unit	/note= "Repeat 3"
DT	01-JUL-1994	(first entry)		FT	exon	5221..5330
DE	ced-3 gene.			FT	exon	/tag= r
KW	C.elegans; cell death; gene; ced-3; inhibition; human; parasite;			FT	mutation	/rpt_type= INVERTED
KW	interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;			FT	mutation	/note= "Repeat 3"
KW	inflammatory response; nematode; diagnosis; myocardial infarction;			FT	mutation	/tag= s
KW	stroke; degenerative disease; traumatic brain injury; hypoxia; pest;			FT	mutation	/number= Exon_5
KW	pathogenic infection; hair loss; cancer; autoreactive antibody; ss.			FT	mutation	5757
OS	Caenorhabditis elegans.			FT	mutation	/tag= t
PH	Key	Location/Qualifiers		FT	mutation	/note= "G>A, from allele n2433"
FT	repeat_unit	1356..1472		FT	intron	5761..5814
FT	repeat_unit	/tag= a		FT	intron	/tag= u
FT	repeat_unit	/rpt_type= INVERTED		FT	intron	/number= Intron_5
FT	repeat_unit	/note= "Inverted w.r.t. repeat at 1490-1614"		FT	intron	5815..5942
FT	repeat_unit	1490..1614		FT	intron	/tag= v
FT	repeat_unit	/tag= b		FT	intron	/number= Exon_6
FT	repeat_unit	/rpt_type= INVERTED		FT	intron	/tag= w
FT	repeat_unit	/note= "Inverted w.r.t. repeat at 1356-1472"		FT	intron	/note= "C>T, from allele n1165"
FT	repeat_unit	2232..2366		FT	intron	5943..6297
FT	repeat_unit	/tag= c		FT	intron	/tag= x
FT	repeat_unit	/number= Exon_1		FT	intron	/number= Intron_6
FT	repeat_unit	2310		FT	intron	6062..6138
FT	repeat_unit	/tag= d		FT	intron	/tag= y
FT	repeat_unit	/note= "C>T, from allele n1040"		FT	intron	/rpt_type= INVERTED
FT	repeat_unit	2367..2429		FT	intron	/tag= z
FT	repeat_unit	/tag= e		FT	intron	/note= "G>A, from allele n717"
FT	repeat_unit	/number= Intron_1		FT	intron	6298..6537
FT	repeat_unit	2430..2575		FT	intron	/tag= aa
FT	repeat_unit	/tag= f		FT	intron	/number= Exon_7
FT	repeat_unit			FT	intron	6322
FT	repeat_unit			FT	intron	/tag= ab
FT	repeat_unit			FT	intron	/note= "C>T, from allele n1949"

FT mutation 6372 /*tag= ac
 FT /*note= "G>A, from allele n1286"
 FT 6434 /*tag= ad
 FT /*note= "C>T, from alleles n1129 and n1164"
 FT 6485 /*tag= ae
 FT /*note= "C>T, from allele n2430"
 FT 6535 /*tag= af
 FT /*note= "G>A, from allele n2426"
 FT 6538..7012 /*tag= ag
 FT /*number= Intron_7
 FT repeat_unit 5567..6625 /*tag= ah
 FT /*rpt_type= INVERTED
 FT repeat_unit /*note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965 /*tag= ai
 FT /*rpt_type= INVERTED
 FT exon /*note= "Inverted w.r.t. repeat 6567-6625"
 FT 7013..7075 /*tag= aj
 FT /*number= Exon_8
 FT mutation 7020 /*tag= ak
 FT /*note= "C>T, from allele n1163"
 FT misc_feature 1..5850 /*tag= al
 FT /*note= "Claim 2, inhibitor fragment"
 FT misc_feature 1..3020 /*tag= am
 FT /*note= "Claim 2, inhibitor fragment"
 FT WO9325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI: 94-007551/01.
 DR P-PSDB; R45262.
 PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Disclosure: Fig 3: 132pp: English.
 CC This sequence encodes the C.elegans cell death gene, ced-3. Fragments
 CC of the amino terminal of the protein encoded by this sequence act as
 CC inhibitors of ced-3. This gene has considerable similarity to human
 CC interleukin-1beta convertase (ICE), which converts pro-interleukin-
 CC 1beta to the active cytokine and is involved in inflammatory response
 CC in humans. The similarity between the two sequences suggests that
 CC inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
 CC nematode ced-3 proteins have an overall amino acid similarity of 28%.
 CC The ced-3 inhibitors may be used for identifying agents which affect
 CC the activity of a gene belonging to the to the ced-3/ICE family of
 CC genes and for diagnosis of diseases characterised by cell death. They
 CC can also be used to develop drugs for treating conditions characterised
 CC by cell deaths such as myocardial infarction, stroke, degenerative
 CC disease, traumatic brain injury, hypoxia, pathogenic infection, or
 CC hair loss, or drugs for reducing the proliferative capacity or size
 CC of a population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 901 gtcgtccttggtatctcaacttgcgcggtttgttttcctggtacactcttcctgtagtc 960
 QY 901 gtcgtccttggtatctcaacttgcgcggtttgttttcctggtacactcttcctgtagtc 960
 Db 961 cactgtctcgtctcaattatctgttagaaatgtgaactgtccagatgggtgactcata 1020
 QY 961 cactgtctcgtctcaattatctgttagaaatgtgaactgtccagatgggtgactcata 1020
 Db 1021 ttgctgtctcgtctcaacttcttcttctcgtcggagcttctcagagcccatataaac 1080
 QY 1021 ttgctgtctcgtctcaacttcttcttctcgtcggagcttctcagagcccatataaac 1080
 Db 1081 tttttttccgcgaaatttgcaataaacccgcgcaaaacttctccaaattgttacgcaa 1140
 QY 1081 tttttttccgcgaaatttgcaataaacccgcgcaaaacttctccaaattgttacgcaa 1140
 Db 1141 tatatacaatccataagaatatctctcaatgtttatgatttcttcgcagcactttctct 1200
 QY 1141 tatatacaatccataagaatatctctcaatgtttatgatttcttcgcagcactttctct 1200
 Db 1201 tctgtgtcgtacatcttatttttataatattccgtctaaattccgatttttgagtatta 1260
 QY 1201 tctgtgtcgtacatcttatttttataatattccgtctaaattccgatttttgagtatta 1260
 Db 1261 attatcgtataaattatcataatagcaccgcaaaactactataaaatgtaaaagctcttt 1320
 QY 1261 attatcgtataaattatcataatagcaccgcaaaactactataaaatgtaaaagctcttt 1320
 Db 1321 taaatcggctcgacattatcgatttaaggaaatcacaaaaattctgagaatgcgtactgcgc 1380
 QY 1321 taaatcggctcgacattatcgatttaaggaaatcacaaaaattctgagaatgcgtactgcgc 1380
 Db 1381 aacatatttgacgtcaaaatctcgttaggggaaacctacagtaattctttaaagctactac 1440
 QY 1381 aacatatttgacgtcaaaatctcgttaggggaaacctacagtaattctttaaagctactac 1440
 Db 1441 tctagcgtctgtcgtattacggggtcgaatttttgaataaatttttttttcggaattt 1500
 QY 1441 tctagcgtctgtcgtattacggggtcgaatttttgaataaatttttttttcggaattt 1500
 Db 1501 tgatacccgtaaaatcgtcacacgctacagtagtcattttaaggattactgtagtctta 1560
 QY 1501 tgatacccgtaaaatcgtcacacgctacagtagtcattttaaggattactgtagtctta 1560
 Db 1561 gctacgagatattttgcgcgcaaaatgatgactgaatacgcattctctgaattttgtgtt 1620
 QY 1561 gctacgagatattttgcgcgcaaaatgatgactgaatacgcattctctgaattttgtgtt 1620
 Db 1621 tccgtaataatttccacaagattttggcattccactttaaagcgccagagattattcca 1680
 QY 1621 tccgtaataatttccacaagattttggcattccactttaaagcgccagagattattcca 1680
 Db 1681 atgggtctcggcagcaaaaagtgttagactttttaaattctccttgattttaattc 1740
 QY 1681 atgggtctcggcagcaaaaagtgttagactttttaaattctccttgattttaattc 1740
 Db 1741 aattactaaaattttcgtgaatttttctgttaaattttttaaatacagttttcttaatt 1800
 QY 1741 aattactaaaattttcgtgaatttttctgttaaattttttaaatacagttttcttaatt 1800
 Db 1801 ttccaggtcgacaaacagaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 1860
 QY 1801 ttccaggtcgacaaacagaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 1860
 Db 1861 taataataacgattttctcattgaaattgtgtttattgtt 1900
 QY 1861 taataataacgattttctcattgaaattgtgtttattgtt 1900

RESULT 3
 ID O64736 standard; DNA; 7653 BP.
 AC O64736;


```
QY 1261 ATTTATCGTAAATTTATCATATAGCACCGGAAACTACTAAAAATGGTAAAAAGCTCCTTT 1320
Dh 1321 taaatcgctcgacattatcgattaaaggaatcacaaaatctgaaatcgctactgcgc 1380
QY 1321 TAAATCGGCTCGACATTATCGTATTAAGGAATCACAATAATCTGAGAAATCGTACTGCGC 1380
Db 1381 aacataattgacgcgaataatctcgtagcgaaactcacagtaattctttaaagtactac 1440
QY 1381 AACATATTGACGGCAAAATATCTCGTAGCGAAACTACAGTAATCTTTAAATGACTAC 1440
Db 1441 ttagcgttgctgctgatttaccgggctcaatttttggaaataatttttttttttcgaattt 1500
QY 1441 TGTAGCGCTTGTGTCGATTTACGGGCTCAATTTTGGAAATAATTTTTCGAATTT 1500
Db 1501 tgataaccgtaaatctcacaacgctcacagtagctgattttaaaggattactgattctta 1560
QY 1501 TGATAACCGTAAATCGTACACGCTACAGTAGTCAATTTAAAGGATTTACTGTAGTTCTA 1560
Db 1561 gctacgacataatttgcgcgcgaataatgactgtaataacgcatctctcgaattttgtgtt 1620
QY 1561 GCTACGAGATATTTGCGCGCAAAATATGACTGTAATACGCAATCTCTGAATTTTGTGTT 1620
Db 1621 tccgtaataattcacaagaatttggcattccactttaaaggcgacagagattattccca 1680
QY 1621 TCCGTAATAATTTTCAAGATTTTGGCATTTCCACTTTAAAGGCGCACAGGATTTATCCA 1680
Db 1681 atgggctgggcacgggaaagatttgatagacttttaaaattctccttgcatctttaaattc 1740
QY 1681 ATGGGCTCGCGCACGCAAAAGTTTGAATAGACTTTTAAATTTCTCTGCAATTTTAAATTC 1740
Db 1741 aattactaaaatttctgtaattttctgttaaatttttaaaatcagttttcttaattt 1800
QY 1741 AATTACTAAAATTTTCGTGAATTTTCTGTTAAATTTTAAATCAGTTTCTTAATATT 1800
Db 1801 ttccaggctgacaaacagaaacaaacacacacacacacacacacacacacacacacacac 1860
QY 1801 TTCCAGGCTGACAAACAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1860
Db 1861 taaaaataacgatttctcattgaaattgtgttatgtt 1900
QY 1861 TAAAAATAACGATTTCTCATTGAAATTTGTGTTTATGTT 1900

RESULT 6
ID Q64744 standard; DNA; 7653 BP.
AC Q64744;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6536A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /tag= f
FT PN W09325685-A.
```

PD 23-DEC-1993.
 PF 14-JUN-1993: U05701
 PK 12-JUN-1992: US-897788.
 PR 20-NOV-1992: US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 PJ wbt: 94-007542/01.
 PK P-PSDB: R53287.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 14: Fig 4; 127pp; English.
 CC The sequences given in 064735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.2%; Score 992; DB 9; Length 7653;
 Best Local Similarity 99.6%; Pred. No. 0.00e+00;
 Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 901 gtctctctgtatcccaactgtcccgctgtgttttttcggtaacacttccgtgatgc 960
 QY 901 GTCGCTCTGGTATCCCACTTGTCGCCGTTGTTTTCGGTACACTCTTCGGTGATGC 960
 Db 961 cactgtctcgtctcaattatcgttttagaattgtgaactgtccagatgggtgactcata 1020
 QY 961 CACGTGTCTCGTCTCAATATATCTTTAGAAATGTGAACGTGCCAGTGGTGACTCATA 1020
 Db 1021 ttactgtgtacaatccactctctttcttcactcggcagctcttacgagccatcaaac 1080
 QY 1021 TTGCTGTCTGTACAACTCACTTTCTTTCTCATCGGCAGCTCTTACGAGCCCATCAAAAC 1080
 Db 1081 ttttttccgcgaaatttgcataaacgcccgaacttctccaaatttttttttttttttt 1140
 QY 1081 TTTTTCGCGCAATTTGCATAAACCAGCCGCAAAACTTCTCCAAATTTGTACGCAA 1140
 Db 1141 tatatacaatccataagaatatcttctcaatgtttatgatgtttcttcgcagcacttctct 1200
 QY 1141 TATATACAAATCCATAAGAAATATCTTCTCAATGTTATGATTTCTTCGACGACTTCTCT 1200
 Db 1201 tcgtgtgtaacaacttatttttataatatttcgcgtaaaatccgatttttgattatta 1260
 QY 1201 TCGTGTGCTAACATCTTATTTTATATATTTTCGGTAAATTCGGATTTTGGATATTA 1260
 Db 1261 atttactgtaaaattatgataatagcaccgaaactactataaaagggttaaaagctccttt 1320
 QY 1261 ATTTATCGTAAATTTATCATATATAGCACCGCAAACTCTTAAATTTGGTAAAGCTCCCTTT 1320
 Db 1321 taaatcggtcgacattatcgatttaagaatcacaaaaattctcgagaatgcgtactgcgc 1380
 QY 1321 TAAATCGGCTCGACATTATCGTATTAAAGAAATCACAAAATCTGAGATGCGTACTGCGC 1380
 Db 1381 aacatatttgacgcaaaatattctcgttagcgaataactacagtaattctttaaatactac 1440
 QY 1381 AACATATTTGACGGCAAAATATCTCGTAGCGAAAACTACAGTAATTTCTTTAAATGACTAC 1440

Db 1441 ttagcgttctgtcgtatcgaattacagggctcaatttttgaaaaataatttttttttcgaattt 1500
 QY 1441 TGTAGGCTTGTGTCGATTTACGGGCTCAATTTTGAAAAATAATTTTTCGAATTT 1500
 Db 1501 tgataaccgtaaatcgtctcaacacgctacagtagctcatttaagagacttagtctcta 1560
 QY 1501 TGATAACCCGTAATCGTCACACACGCTACAGTAGTCAATTTAAAGGATTACTGTAGTTCTA 1560
 Db 1561 gtagcagcatattttgcgcgcacaaatagactgaactgaatcgcattctcgaattttgtgtt 1620
 QY 1561 GTACAGAGATATTTGCGCGCCAAATAGACTGTAATACGCATTTCTGAAATTTGTGTT 1620
 Db 1621 tccgtaataatttcacaagattttggcattccactttaaaggcgacagggatttattcca 1680
 QY 1621 TCCGTAATAATTTCAAGATTTTGGCATTTCCACTTTAAAGCGCACAGGATTTATTTCCA 1680
 Db 1681 atgggtctggcagcaggaagaaagtgtgacacttttaaatctccttgacattttaattc 1740
 QY 1681 ATGGGTCTCGGCACGCAAAAGTTTGTAGACTTTTAAATTTCTCCTTGCATTTTAATTC 1740
 Db 1741 aattactaaaaattttcgtgaatttttctgtttaaatttttaaaatcagttttctaaatt 1800
 QY 1741 AATTACTAAAATTTTCGTGAATTTTCTGTTAAATTTTAAATTCAGTTTCTTAATTT 1800
 Db 1801 ttccaggctgacaacagaaac 1860
 QY 1801 TTCCAGGCTGACAAACAGAAACAAAAACACAAACAAACAAACATTTTAAAAATCAGTTTTCAAAT 1860
 Db 1861 taaaaataacgatttctcattgaaattgtgtttatgttt 1900
 QY 1861 TAAAAATAACGATTTCTCATTGAAATTTGTGTTTATGTT 1900

RESULT 7
 ID 064739 standard; DNA; 7653 BP.
 AC 064739; 1994 (first entry)
 DT 23-JUN-1994
 DE ced-3 (G6297A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT repeat_unit 1356..1472
 FT /*tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT repeat_unit 1490..1614
 FT /*tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT exon 2232..2366
 FT /*tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /*tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /*tag= e
 FT /number= Exon_2
 FT 2576..2853
 FT /*tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /*tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /*tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /*tag= i
 FT repeat_unit
 FT

```

FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit
FT 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit
FT 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit
FT 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon
FT 4303..4634
FT /*tag= m
FT /number= Exon_4
FT 4635..5546
FT /*tag= n
FT /number= Intron_4
FT 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit
FT 5221..5330
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon
FT 5547..5760
FT /*tag= q
FT /number= Exon_5
FT 5761..5814
FT /*tag= r
FT /number= Intron_5
FT exon
FT 5815..5942
FT /*tag= s
FT /number= Exon_6
FT 5943..6297
FT /*tag= t
FT /number= Intron_6
FT mutation
FT 5757
FT /*tag= u
FT /note= "C>T, from allele n717"
FT repeat_region
FT 6062..6138
FT /*tag= v
FT /rpt_type= INVERTED
FT exon
FT 6298..6537
FT /*tag= w
FT /number= Exon_7
FT 6538..7012
FT /*tag= x
FT /number= Intron_7
FT repeat_unit
FT 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit
FT 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon
FT 7012..7075
FT /*tag= aa
FT /number= Exon_8
FT
FT W09325685-A.
FT 23-DEC-1993.
FT PD
FT 14-JUN-1993: U05701.
FT PR
FT 12-JUN-1992: US-897788.
FT PR
FT 20-NOV-1992: US-979638.
FT PA
FT (NASI ) MASSACHUSETTS INST TECHNOLOGY.
FT PI
FT Horvitz HR, Shaham S, Yuan J;
FT WPI; 94-007542/01.
FT DR
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms

```

```

PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myob. The ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
CC Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
SQ
Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 901 gtcgtccttggtatcctcaactgtcccggtttgttttcggtacactttccggtgagtc 960
QY |||||
901 GTCGCTCTTGGTATCTCAACTGTGCCGGTTTGTTCGGTACACTTTCGGTGATGC 960
Db 961 cactgtctcgtctcaatattgtagaataatggaactgccagtgagtcactcata 1020
QY |||||
961 CACCTGTCTCCGTCCTCAATATCGTTAGAAATGTGAACCTCCAGATGGGTGACTATA 1020
Db 1021 ttgctgctgctacaatccactttcttctcgcagctcttacgagcccatcataaac 1080
QY |||||
1021 TTGCTGCTGCTACAATCCACTTTCCTTCATCGCGCAGCTTACGAGCCCATATAAAC 1080
Db 1081 tttttttccgcgaaatttgcataaaacggcccaaaattttcccaatttgtacgaaa 1140
QY |||||
1081 TTTTTCCTCCGGAATTTGCAATAAAGCGGCCAAAACCTTCTCCAAAATTTGTACGCAA 1140
Db 1141 tatatacaatccaataagaatatcttctcaattgttatgtattcttcgcgagcacttctct 1200
QY |||||
1141 TATATACAATCCATAAGAATATCTTCTCAATGTTTATGATTTCTTCGAGCAGCTTCTCT 1200
Db 1201 tcgtgctacaactcttatttttataatatttccgctaaaaattccgatttttgagtatta 1260
QY |||||
1201 TCGTGTGCTAACATCTTATTTTATAATAATTTCCGCTAAAAATCCGATTTTTCGATATTA 1260
Db 1261 attatcgtaaaattatgataatagaccgaaaactactaaaaatggtaaaactctcttt 1320
QY |||||
1261 ATTTATCGTAAAAATTATCATATAATAGCACCGAAAACACTACTAAAAATGGTAAAGCTCCTTT 1320
Db 1321 taatcggtcgcacattatcgattatgaaggaatcacaaaaattctgagaatcgctactgcgc 1380
QY |||||
1321 TAAATCGGCTCGACATATTCGTATTAGGAATCACAAAATTCAGAGATCGGTACTGCGC 1380
Db 1381 aacatattgacggcgaataatctcgtagcgaatactacagtaatttcttaaatgactac 1440
QY |||||
1381 AACATAATTTGACGGCAAAATATCTCGTAGCGAAAACFACAGTAATTTCTTTAAATGACTAC 1440
Db 1441 tgtagcgttggtcgtgatttacggggtcaatttttgaataaatttttttttcgaattt 1500
QY |||||
1441 TGTAGCGCTTGTGTCGATTTACGGGCTCAATTTTGAATAATAATTTTTCGAAATTT 1500
Db 1501 tgataacccgttaaatcgtcaacgcgtacagtagtacttaaaaggattactgtagtcta 1560
QY |||||
1501 TGATAACCCGTAATAATCGTCACACCTACAGTACTCATTTAAAGGATTACTGTAGTCTA 1560
Db 1561 gctacgacataattttgcgcgcaaatatgactgtaatacgcattctctcgaattttgtgt 1620

```


[illegible]

RESULT	8	
ID	Q64745	standard; DNA; 7653 BP.
AC	Q64745;	
DT	23-JUN-1994	(first entry)
DE	ced-3	(C7020T) gene.
DE	ced-3	(C7020T) gene.
KW	<i>C. elegans</i> ;	ced-4; ced-3; mutant; transcriptional regulation;
KW	embryogenesis;	cell death; hydrophilic; transmembrane; region;
KW	hydrophobic;	mutation; amino acid; substitution; RNA splicing;
KW	protein synthesis;	null phenotype; calcium-binding domain; ss.
OS	<i>Caenorhabditis elegans</i> .	

FT	repeat_unit	Location/Qualifiers
FT		1356..1472
FT		/*tag= a
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t repeat at 1490-1614"
FT	repeat_unit	1490..1614
FT		/*tag= b
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 1356-1472"
FT	exon	2232..2356
FT		/*tag= c
FT		/number= Exon_1
FT	intron	2367..2429
FT		/*tag= d
FT		/number= Intron_1
FT	exon	2450..2575
FT		/*tag= e
FT		/number= Exon_2
FT	intron	2576..2853
FT		/*tag= f
FT		/number= Intron_2
FT	exon	2854..3107
FT		/*tag= g
FT		/number= Exon_3
FT	intron	3108..4302
FT		/*tag= h
FT		/number= Intron_3
FT	repeat_unit	3126..3243
FT		/*tag= i
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3329-3396"
FT	repeat_unit	3329..3396
FT		/*tag= j
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3126-3243"
FT	repeat_unit	3487..3759
FT		/*tag= k
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3782-4070"
FT		

FT	repeat_unit	3782..4070	
FT		/*tag= 1	
FT		/rpt_type= INVERTED	
FT		/note="Inverted w.r.t. repeat 3487-3759"	
FT	exon	4303..4634	
FT		/*tag= m	
FT		/number= Exon_4	
FT	intron	4635..5546	
FT		/*tag= n	
FT		/number= Intron_4	
FT	repeat_unit	4688..4719	
FT		/*tag= o	
FT		/rpt_type= INVERTED	
FT		/note="Repeat 3"	
FT	repeat_unit	5221..5330	
FT		/*tag= p	
FT		/rpt_type= INVERTED	
FT		/note="Repeat 3"	
FT	exon	5547..5760	
FT		/*tag= q	
FT		/number= Exon_5	
FT	intron	5761..5814	
FT		/*tag= r	
FT		/number= Intron_5	
FT	exon	5815..5942	
FT		/*tag= s	
FT		/number= Exon_6	
FT	intron	5943..6297	
FT		/*tag= t	
FT		/number= Intron_6	
FT	repeat_region	6062..6138	
FT		/*tag= u	
FT		/rpt_type= INVERTED	
FT	exon	6298..6537	
FT		/*tag= v	
FT		/number= Exon_7	
FT	intron	6538..7012	
FT		/*tag= w	
FT		/number= Intron_7	
FT	repeat_unit	6567..6625	
FT		/*tag= x	
FT		/rpt_type= INVERTED	
FT		/note="Inverted w.r.t. repeat at 6905-6965"	
FT	repeat_unit	6905..6965	
FT		/*tag= y	
FT		/rpt_type= INVERTED	
FT	exon	7012..7075	
FT		/*tag= z	
FT		/number= Exon_8	
FT	mutation	7076	
FT		/*tag= w	
FT		/note="C>T, from allele n1163"	
PN	W09325685-A.		
PD	23-DEC-1993.		
PF	14-JUN-1993; 005701.		
PR	12-JUN-1992; US-897788.		
PR	20-NOV-1992; US-979638.		
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.		
PI	Horvitz HR, Shahan S, Yuan J;		
DR	WPI: 94-007542/01.		
DR	P-PSDB; R53288.		
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to		
PT	develop agents to increase or prevent cell death in organisms		
PS	Claim 14; Fig 4; 127pp; English.		
CC	The sequences given in 064735-45 represent mutations of the C		
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcr		
CC	was most abundant in embryos, but was also detected in larvae		
CC	adults, suggesting that ced-3 is expressed not only in cells		
CC	cell death. The four largest introns as well as sequences 5'		
CC	start codon contain repetitive elements, some of which have b		
CC	characterised in non-coding regions of other C. elegans genes		
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 103 amino ac		

Query Match	99.2%	Score 992;	DB 9;	Length 7653;
Best Local Similarity	99.6%;	Pred. No. 0.00e+00;		
Matches	996;	Conservative	0;	Indels 0; Gaps 0;
Db	901	gtcgctctgggtatctcaactgtcccggtttgtttccggtacactcttcggtgatgc	960	
Qy	901			
Db	961	cacctgtccggtctcaattatcgtttagaagtgaactgccagatgggtgactcata	1020	
Qy	961	CACCTGTCCTCGTCAATATATCGTTTGAAGATGTGAACGTGCCAGATGGTGACTCATA	1020	
Db	1021	ttgctgtgtacaaatccacttttctttctcatcggcagctcttcagagcccatataaac	1080	
Qy	1021	TTGCTGTGTCTACAATCCACTTCTTTTCTCATCGGCAGTCTTACGAGCCCATCAATAAC	1080	
Db	1081	ttttttccgcgaatttgcaataaacgcgcgaataaactttccaaaattgttacgcaa	1140	
Qy	1081	TTTTTTTTCCGCGAATTTGCAATAAACCGGCCAAAAACTTTCCTCAAAATGTTTACGCAA	1140	
Db	1141	tataacaatccataagaatatctctcctcaatgtttatgattcttcgcagcaactttctct	1200	
Qy	1141	TATATACAATCCATAAGAAATATCTTCTCAATGTTTATGATTTCTTCGCAGCACTTCTCT	1200	
Db	1201	tcgtgtcctaacaattttttataataatttcgcgtaaaaattccgatttttgagtatta	1260	
Qy	1201	TCGTGTCTCAACATCTATTTTTTATAATATTTCCGCTAAAAATCCGATTTTGTAGTATTA	1260	
Db	1261	attctcgtaaaattatgataatagcaccgcgaataactaaaaatggtaaaagctcttt	1320	
Qy	1261	ATTTATCGTAAAAATTTATATATAGCACCGCAAAACTACTAAAAATGGTAAAGCTCCTTT	1320	
Db	1321	taaatcggtcgcacattatcgattaaaggaatccaaaaattctgagaatcgctactgcgc	1380	
Qy	1321	TAAATCGGCTCGACATTATCGTATTAAAGGAATCACAAAATTTCTGAGAATGCGTACTGCGC	1380	
Db	1381	aacataattgacggcaaaaatctctcfagcgaataactcacagtaattctttaaatgaactac	1440	
Qy	1381	AACATATTGACGCGCAAAATATCTGTCAGCGAAAACACTACAGTAATTTCTTTAAATGACTAC	1440	
Db	1441	tgtacgcctgtgtcgattttacgggtcgaattttgaaaaataatttttttcgcaatt	1500	
Qy	1441	TGTAGCGCTGTGTGCGATTTTACGGGCTCAATTTTTTGAAAAATAATTTTTTTTTTCGAATTT	1500	
Db	1501	tgataaacccgttaaatcgtcacacgcgtacagtagtcattttaaaggattactgtagtctta	1560	
Qy	1501	TGATAACCCGTAAATCGTCACACGCTACAGTAGTCATTTAAAGGATTACTGTAGTTCTTA	1560	
Db	1561	gctacgacataattttgcgcgcgaataatgactgtataacgactctctggaatttttggtt	1620	
Qy	1561	GCTACGAGATTTTTTTCGCGCCCAAAATGACTGTATAGCAATCTCTGSAATTTTGTGTT	1620	
Db	1621	lccgtataaatttcacaagattttgcattccactttaaagcgcacaggaattattccca	1680	
Qy	1621	TCCGTAATAATTTTTCACAGATTTTGGCATTTCCACTTTTAAAGCGGCACAGGAAATTTATTC	1680	
Db	1681	atgggctctgggcgcgcaaaagtttgatagacttttaaatctctctgcatttttaattc	1740	

FT /number= Intron_4
 FT 4688..4719
 FT /tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /tag= q
 FT /number= Exon_5
 FT 5761..5814
 FT /tag= r
 FT /number= Intron_5
 FT 5815..5942
 FT /tag= s
 FT /number= Exon_6
 FT 5943..6297
 FT /tag= t
 FT /number= Intron_6
 FT 6062..6138
 FT /tag= u
 FT /rpt_type= INVERTED
 FT 6298..6537
 FT /tag= v
 FT /number= Exon_7
 FT 6434
 FT /tag= w
 FT /note= "C>T, from allele n1129 and n1164"
 FT 6538..7012
 FT /tag= x
 FT /number= Intron_7
 FT 6567..6625
 FT /tag= y
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /tag= z
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7012..7075
 FT /tag= aa
 FT /number= Exon_8
 FT
 PN W0325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; US-897788.
 PR 12-JUN-1992; US-979638.
 PR 20-NOV-1992; US-979638.
 PI (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PA Horvitz HR, Shaham S, Yuan J;
 DR WPI; 94-007542/01.
 DR P-PSDB; R53285.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 14; Fig 4; 127pp; English
 CC The sequences given in Q64735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene

CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
 Query Match 99.2%; Score 992; DB 9; Length 7653;
 Best Local Similarity 99.6%; Pred. No. 0.00e+00;
 Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Db 901 gtcgctctggtatctcctcaactgtcccggttttggcttcgtacactcttcogtgatgc 960
 QY 901 gtcgctctggtatctcctcaactgtcccggttttggcttcgtacactcttcogtgatgc 960
 Db 961 cactgtctctcgtctcaattatcggttagaactgtgaactgccagatgccagtgactcata 1020
 QY 961 cactgtctctcgtctcaattatcggttagaactgtgaactgccagatgccagtgactcata 1020
 Db 1021 ttgctgtctacacatccacttttcttcctatcgccagctcttaccagcccatcaaac 1080
 QY 1021 ttgctgtctacacatccacttttcttcctatcgccagctcttaccagcccatcaaac 1080
 Db 1081 tttttttccgggaaatttgcataaaacggcgaacaaatttctccaaattttacgaaa 1140
 QY 1081 tttttttccgggaaatttgcataaaacggcgaacaaatttctccaaattttacgaaa 1140
 Db 1141 tatatacaatccataagaatatctctcctcaattgtttatgattctcttcgcagcactctct 1200
 QY 1141 tatatacaatccataagaatatctctcctcaattgtttatgattctcttcgcagcactctct 1200
 Db 1201 tcgtgtctacacatcttttttataataatttcgcgtataatttcgcgtattttgagtatta 1260
 QY 1201 tcgtgtctacacatcttttttataataatttcgcgtataatttcgcgtattttgagtatta 1260
 Db 1261 atttctgtaaaattatgataatagcagcgaacaaactactataaaatggtaaaagctccitt 1320
 QY 1261 atttctgtaaaattatgataatagcagcgaacaaactactataaaatggtaaaagctccitt 1320
 Db 1321 taaatcggtctcgacattatcgattaaaggaatcacaaaattctgagaatggctactgcgc 1380
 QY 1321 taaatcggtctcgacattatcgattaaaggaatcacaaaattctgagaatggctactgcgc 1380
 Db 1381 aacatatattgacggcaaaaatactctgtagcgaaaactacagtaattctttaaatactac 1440
 QY 1381 aacatatattgacggcaaaaatactctgtagcgaaaactacagtaattctttaaatactac 1440
 Db 1441 tctagcgtctgtcgatttcaggtctcaatttttgaaataatttttttttttttttttttt 1500
 QY 1441 tctagcgtctgtcgatttcaggtctcaatttttgaaataatttttttttttttttttttt 1500
 Db 1501 tgataaccggttaaatctgcacacgctacagtagtctatttaagagattactgttagtcta 1560
 QY 1501 tgataaccggttaaatctgcacacgctacagtagtctatttaagagattactgttagtcta 1560
 Db 1561 gctacgacatatttgcgcgcacaaataatgactgtaataatgcattctctgaattttgtgtt 1620
 QY 1561 gctacgacatatttgcgcgcacaaataatgactgtaataatgcattctctgaattttgtgtt 1620
 Db 1621 tccgtaataatttcacagattttgcattccactttaaagcgacagattatttcca 1680
 QY 1621 tccgtaataatttcacagattttgcattccactttaaagcgacagattatttcca 1680
 Db 1681 atgggtctggggcacggaaaagtttgatagacttttaaatctccttcgatttttaatttc 1740
 QY 1681 atgggtctggggcacggaaaagtttgatagacttttaaatctccttcgatttttaatttc 1740
 Db 1741 aattactaaaatttctcgtaattttcttgtaaaattttttaaataatgcattttcctaatt 1800
 QY 1741 aattactaaaatttctcgtaattttcttgtaaaattttttaaataatgcattttcctaatt 1800
 Db 1801 ttccaggctgacaaacagaaaacaaacacacacacacacacacacacacacacacacacacac 1860
 QY 1801 ttccaggctgacaaacagaaaacaaacacacacacacacacacacacacacacacacacacac 1860
 Db 1801 ttccaggctgacaaacagaaaacaaacacacacacacacacacacacacacacacacacacac 1860
 QY 1801 ttccaggctgacaaacagaaaacaaacacacacacacacacacacacacacacacacacacac 1860

```

Db 1861 taaaataacgatttcattgaaattgtttatatt 1900
    |||||||
QY 1861 TAAAAATACGATTTCATTGAAATGTGTTTATGTT 1900

RESULT 10
ID Q64743 standard; DNA; 7653 BP.
AC Q64743;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6485f) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /tag= g
FT /number= Exon_3
FT intron 3108..4302
FT /tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /tag= m
FT /number= Exon_4
FT intron 4635..5546
FT /tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"

```

```

FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT repeat_region 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT mutation 6485
FT /tag= w
FT /note= "C>T, from allele n2430"
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT WO9325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shahan S, Yuan J;
FT WPI; 94-007542/01.
FT P-PSDB; R53286.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT start codon contain repetitive elements, some of which have been
FT characterised in non-coding regions of other C. elegans genes, such
FT as lem-1, lin-12 and myob. The Ced-3 protein is 503 amino acids in
FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT region that might be a transmembrane region. One region of Ced-3 is
FT very rich in serine. It is thought that this region is involved in
FT protein-protein interactions, similar to acid blobs in transcription
FT factors. Of the mutations which occur within the ced-3 gene, eight of
FT the mutations are missense mutations, two are nonsense mutations and
FT two are putative splicing mutations. These mutations establish the
FT null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT function is not essential for viability. The ced-3 and ced-4 gene
FT products may be used to develop agents for treating conditions
FT characterised by cell deaths, such as myocardial infarction, stroke,
FT degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT infection, aging or hair loss.
FT Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
FT
FT Query Match 99.2%; Score 992; DB 9; Length 7653;
FT Best Local Similarity 99.6%; Pred. No. 0.00e+00;
FT Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

Db 901 gtcgtcttggtatctcctcaactgtcccggtttgttttcggtacactcttcggtgacgc 960
 QY 901 gtcgtcttggtatctcctcaactgtcccggtttgttttcggtacactcttcggtgacgc 960
 Db 961 caccctctccgtctcaaatatcggtttagaagaatggaactgccagatgggtgactcata 1020
 QY 961 caccctctccgtctcaaatatcggtttagaagaatggaactgccagatgggtgactcata 1020
 Db 1021 ttgctgctgacacatccactcttcttctcgcagctcttctcagagcccatcataaac 1080
 QY 1021 ttgctgctgacacatccactcttcttctcgcagctcttctcagagcccatcataaac 1080
 Db 1081 tttttttccgcggaaatttgcataaacacgcggcaaaaactcttccaaattgttacgcaa 1140
 QY 1081 tttttttccgcggaaatttgcataaacacgcggcaaaaactcttccaaattgttacgcaa 1140
 Db 1141 tatatacaatccataagaatatctctcctcaatgtttatgatattcttcgcagcacttctct 1200
 QY 1141 tatatacaatccataagaatatctctcctcaatgtttatgatattcttcgcagcacttctct 1200
 Db 1201 tcgtgtgtaacatcttattttataataatttcgcgtaaaattccgatttttggagtatta 1260
 QY 1201 tcgtgtgtaacatcttattttataataatttcgcgtaaaattccgatttttggagtatta 1260
 Db 1261 attatcgtaaaattatgataatagcaccgaaactactcaaaaatggtaaaagctccctt 1320
 QY 1261 attatcgtaaaattatgataatagcaccgaaactactcaaaaatggtaaaagctccctt 1320
 Db 1321 taatcggtcgacattatcgtaattgaagaatcacaaaattctgagaatgcgtactgcgc 1380
 QY 1321 taatcggtcgacattatcgtaattgaagaatcacaaaattctgagaatgcgtactgcgc 1380
 Db 1381 aacatatgtgacgcaaaatattctctagcgaatactcagtaattctttaaataactact 1440
 QY 1381 aacatatgtgacgcaaaatattctctagcgaatactcagtaattctttaaataactact 1440
 Db 1441 tctagcgttctgctgattacgggttcgaattttgaaataaatttttttttttcgaattt 1500
 QY 1441 tctagcgttctgctgattacgggttcgaattttgaaataaatttttttttttcgaattt 1500
 Db 1501 tgatacccgtaaatcgacacacgtacagtagtcatcttaaaggattactgtactcta 1560
 QY 1501 tgatacccgtaaatcgacacacgtacagtagtcatcttaaaggattactgtactcta 1560
 Db 1561 gctacgacatattttgctgcgcgaataatgactgtatacgcattctctgaaattttgtgtt 1620
 QY 1561 gctacgagatattttgctgcgcgaataatgactgtatacgcattctctgaaattttgtgtt 1620
 Db 1621 tccgtaataatttcaagaattttggtcattccacttcaaaaggcagcagattttatcca 1680
 QY 1621 tccgtaataatttcaagaattttggtcattccacttcaaaaggcagcagattttatcca 1680
 Db 1681 atgggtctggcgcgcaaaaagttgatgacttttaaatctccttgcatctttaaattc 1740
 QY 1681 atgggtctggcgcgcaaaaagttgatgacttttaaatctccttgcatctttaaattc 1740
 Db 1741 aattactaaaattttcgtggaattttctgtttaaatttttaaaatcagttttctaatt 1800
 QY 1741 aattactaaaattttcgtggaattttctgtttaaatttttaaaatcagttttctaatt 1800
 Db 1801 ttcaggctgacaaaacagaacaaaacacacacacatttttaaaatcagttttcacaat 1860
 QY 1801 ttcaggctgacaaaacagaacaaaacacacacacatttttaaaatcagttttcacaat 1860
 Db 1861 taataataacgattttcattgaaattgtgttttatgtt 1900
 QY 1861 taataataacgattttcattgaaattgtgttttatgtt 1900
 Db 1901 taaaaataacgattttcattgaaattgtgttttatgtt 1900
 QY 1901 taaaaataacgattttcattgaaattgtgttttatgtt 1900

RESULT 11
 ID Q64740 standard: DNA; 7653 BP.
 AC Q64740;

23-JUN-1994 (first entry)
 DE ced-3 (C6322T) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 embryogenesis; cell death; hydrophilic; transmembrane; region;
 hydrophobic; mutation; amino acid; substitution; RNA splicing;
 protein synthesis; null phenotype; calcium-binding domain; ss.
 KW Caenorhabditis elegans.
 OS Caenorhabditis elegans.
 PH Key Location/Qualifiers
 FT repeat_unit 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT repeat_unit 1490..1614
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT exon 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT intron 2367..2429
 FT /tag= d
 FT /number= Intron_1
 FT exon 2450..2575
 FT /tag= e
 FT /number= Exon_2
 FT intron 2576..2853
 FT /tag= f
 FT /number= Intron_2
 FT exon 2854..3107
 FT /tag= g
 FT /number= Exon_3
 FT intron 3108..4302
 FT /tag= h
 FT /number= Intron_3
 FT repeat_unit 3126..3243
 FT /tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT repeat_unit 3329..3396
 FT /tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT repeat_unit 3487..3759
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT repeat_unit 3782..4070
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT exon 4303..4634
 FT /tag= m
 FT /number= Exon_4
 FT intron 4635..5546
 FT /tag= n
 FT /number= Intron_4
 FT repeat_unit 4688..4719
 FT /tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT repeat_unit 5221..5330
 FT /tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT exon 5547..5760
 FT /tag= q
 FT /number= Exon_5
 FT intron 5761..5814
 FT /tag= r
 FT /number= Intron_5
 FT exon 5815..5942
 FT /tag= s
 FT /number= Exon_6

intron 5943..6297
 /tag= t
 /number= Intron_6
 6062..6138
 /tag= u
 /rpt_type= INVERTED
 6298..6537
 /tag= v
 /number= Exon_7
 6322
 /tag= w
 /note= "C>T, from allele n1949"
 6538..7012
 /tag= x
 /number= Intron_7
 6567..6625
 /tag= y
 /rpt_type= INVERTED
 /note= "Inverted w.r.t. repeat at 6905-6965"
 6905..6965
 /tag= z
 /rpt_type= INVERTED
 /note= "Inverted w.r.t. repeat 6567-6625"
 7012..7075
 /tag= aa
 /number= Exon_8

W09325685-A.

23-DEC-1993.

14-JUN-1993; U05701.

12-JUN-1992; US-897788.

20-NOV-1992; US-979638.

(MASI) MASSACHUSETTS INST TECHNOLOGY.

Horvitz HR, Shaham S, Yuan J;

P-PSDB; R53283.

Isolated C elegans cell death genes ced-3 and ced-4 - used to

develop agents to increase or prevent cell death in organisms

Claim 14; Fig 4; 127pp; English.

The sequences given in 04735-45 represent mutations of the C. elegans ced-3 gene. A 2.8 kb region was identified as the ced-3 transcript and as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is highly hydrophilic with no significant hydrophobic adults, suggesting that ced-3 is expressed not only in cells undergoing cell death. The four largest introns as well as sequences 5' of the start codon contain repetitive elements, some of which have been characterized in non-coding regions of other C. elegans genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is highly hydrophilic with no significant hydrophobic region that might be a transmembrane region. One region of Ced-3 is very rich in serine. It is thought that this region is involved in protein-protein interactions, similar to acid blobs in transcription factors. Of the mutations which occur within the ced-3 gene, eight of the mutations are missense mutations, two are nonsense mutations and two are putative splicing mutations. These mutations establish the null phenotype of the ced-3 gene, confirming that ced-3, like ced-4, function is not essential for viability. The ced-3 and ced-4 gene products may be used to develop agents for treating conditions characterized by cell deaths, such as myocardial infarction, stroke, degenerative disease, traumatic brain injury, hypoxia, pathogenic infection, aging or hair loss.

Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.28; Score 992; DB 9; Length 7653;

Best Local Similarity 99.68; Pred. No. 0.00e+00; Mismatches 4; Indels 0; Gaps 0; Matches 996; Conservative 0;

901 gtgcgtctgtatcccaactgtccggtttgttttgcgtacacttccgtgatgc 960
 |||||
 901 GTGCTCTTGGTATCTCACTTGTCCCGGTTTGTTCGGTACACTCTTCGCTGATGC 960
 |||||
 961 caactgtctcgtctcaattatcggttagaaatgtgaactgtccagatgggtgactcata 1020
 |||||
 961 CACCTGTCCTCGTCTCAATATCGTTTAGAAATGTGAACGTGCCAGATGGTGACTCATA 1020

Db 1021 ttgctgtgtacataccactttcttctcatcgcgagcttctacgagcccatcataaac 1080
 |||||
 Qy 1021 TTGCTGTGTACATAACCACTTCTTTTCATCGGAGTCITACGAGCCATCATAAAC 1080
 |||||
 Db 1081 ttttttccgcgaaattgtcaataaacaccgcccacaaactttctccaaattgttacgcaa 1140
 |||||
 Qy 1081 TTTTTCGCGGAAATTGCAATTAACCGCCCAAAACCTTCTCCAAATTTGTACGCA 1140
 |||||
 Db 1141 tatacaatccataagaatatcttctcaatgtttatgtattcttcgcgacactttctct 1200
 |||||
 Qy 1141 TATATACATCCATAGATATCTTCTCAATGTTTATGATTTCTTCGACACTTTCTCT 1200
 |||||
 Db 1201 tcgtgtcctaactcttattttataataatttcgcgtataaattccatatttcagattta 1260
 |||||
 Qy 1201 TCGTGTGCTAACATCTTATTATTAATAATTCCGCTAAATTTCCGATTTTGGAGTATA 1260
 |||||
 Db 1261 attatcgtaaaaattatgataatagcaccgaaaactactaaaaatggttaaaagctctct 1320
 |||||
 Qy 1261 ATTATCGTAAATTTATCATATAGCACCAGAAAACCTACTAAAAATGGTAAAGCTCCTTT 1320
 |||||
 Db 1321 taatcgcgtcgacattatcgatttaaggaatcacaaaattctgagaatcgctactgcgc 1380
 |||||
 Qy 1321 TAAATCGGCTCGACATTTATCGTATTAAGGAATCACAAAATCTTGAGAAATCGTACTGCGC 1380
 |||||
 Db 1381 aacatattgacggcgaataatctcgtagcgaaaactcacagtaattctttaaagactac 1440
 |||||
 Qy 1381 AACATATTGACGGCAAAATATCTGTAGGCAAACTACAGTAATTTCTTTAATGACTAC 1440
 |||||
 Db 1441 tgtagcgttgtgtcgatttacggggtcctcaatttttgaataaatttttttttcgaattt 1500
 |||||
 Qy 1441 TGTAGCGCTTGTGTCGATTACGGGCTCAATTTTGAATAATAATTTTTCGAATTT 1500
 |||||
 Db 1501 tgataacccgtataatcgtacacacgcgtacgtagctctttaaaggattactgtagtcta 1560
 |||||
 Qy 1501 TGATAACCCGTAAATCGTACACGCTACAGTAGTCAATTTAAAGGATTAAGTGTCTA 1560
 |||||
 Db 1561 gctacgacataatttgcgcgcccataatgactgtaacgcatctctgaattttgtgtt 1620
 |||||
 Qy 1561 GCTACGAGATATTTCGCGGCAAAATATGACTGTAATACGCAATCTCTGAATTTGTGTT 1620
 |||||
 Db 1621 tccgtataatttcacaaagattttggcattccactttaaaggcgacagattatttcca 1680
 |||||
 Qy 1621 TCCGTAATAATTTCACAAAGATTTCGATTCCTCCATTTAAAGGCGCACAGGATTTATCCA 1680
 |||||
 Db 1681 atgggtctgcgcgagaaagtgtgtagacttttaaatctcctgcgactttttaaattc 1740
 |||||
 Qy 1681 ATGGGTCTCGCGCAGCAAAAGATTGTAGACTTTTAAATTCCTCTGCAATTTTAAATTC 1740
 |||||
 Db 1741 aattactaaaaatttctgtaattttctgttaaatttttaaaatcagttttctaatatt 1800
 |||||
 Qy 1741 AATTACTAAAAATTTTCGTGAATTTTCTGTAAATTTTAAATTCAGTTTCTTAATATT 1800
 |||||
 Db 1801 ttccaggctgacaaacagaaac 1860
 |||||
 Qy 1801 TTCCAGGCTGACAAACAGAAACACAAACACAAACACAAACATTTTAAATAATCAGTTTCAAT 1860
 |||||
 Db 1861 taaaaataacgatttctcattgaaattgtgtttatgtt 1900
 |||||
 Qy 1861 TAAAAATAACGATTCTCAATGAAATTTGTGTTTATGTT 1900
 |||||

RESULT 12

ID Q54666 standard; DNA; 7653 BP.

AC Q54666;

DE 23-JUN-1994 (first entry)

DE ced-3 gene.

KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;

KW embryogenesis; cell death; hydrophilic; transmembrane; region;

KW hydrophobic; mutation; amino acid; substitution; RNA splicing;

KW protein synthesis; null phenotype; calcium-binding domain; ss.

OS Caenorhabditis elegans.

PH Key Location/Qualifiers

FT repeat_unit 1356..1472

```

FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT repeat_unit
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT exon
FT FT 2232..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2310
FT FT /*tag= d
FT FT /note= "C>T, from allele n1040"
FT FT intron
FT FT 2367..2429
FT FT /*tag= e
FT FT /number= Intron_1
FT FT exon
FT FT 2430..2575
FT FT /*tag= f
FT FT /number= Exon_2
FT FT 2487
FT FT /*tag= g
FT FT /note= "G>A, from allele n718"
FT FT intron
FT FT 2576..2853
FT FT /*tag= h
FT FT /number= Intron_2
FT FT exon
FT FT 2854..3107
FT FT /*tag= i
FT FT /number= Exon_3
FT FT intron
FT FT 3108..4302
FT FT /*tag= j
FT FT /number= Intron_3
FT FT repeat_unit
FT FT 3126..3243
FT FT /*tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT repeat_unit
FT FT 3329..3396
FT FT /*tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT repeat_unit
FT FT 3487..3759
FT FT /*tag= m
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT repeat_unit
FT FT 3782..4070
FT FT /*tag= n
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT exon
FT FT 4303..4634
FT FT /*tag= o
FT FT /number= Exon_4
FT FT intron
FT FT 4635..5546
FT FT /*tag= p
FT FT /number= Intron_4
FT FT repeat_unit
FT FT 4688..4719
FT FT /*tag= q
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT repeat_unit
FT FT 5221..5330
FT FT /*tag= r
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT exon
FT FT 5547..5760
FT FT /*tag= s
FT FT /number= Exon_5
FT FT 5757
FT FT /*tag= t
FT FT /note= "G>A, from allele n2433"
FT FT intron
FT FT 5761..5814
FT FT /*tag= u
FT FT /number= Intron_5
FT FT exon
FT FT 5815..5942
FT FT /*tag= v
FT FT /number= Exon_6

```

```

FT mutation
FT 5940
FT /*tag= w
FT /note= "C>T, from allele n1165"
FT intron
FT 5943..6297
FT /*tag= x
FT /number= Intron_6
FT repeat_region
FT 6062..6138
FT /*tag= y
FT /rpt_type= INVERTED
FT mutation
FT 6297
FT /*tag= z
FT /note= "G>A, from allele n717"
FT exon
FT 6298..6537
FT /*tag= aa
FT /number= Exon_7
FT mutation
FT 6322
FT /*tag= ab
FT /note= "C>T, from allele n1949"
FT mutation
FT 6372
FT /*tag= ac
FT /note= "G>A, from allele n1286"
FT mutation
FT 6434
FT /*tag= ad
FT /note= "C>T, from alleles n1129 and n1164"
FT mutation
FT 6485
FT /*tag= ae
FT /note= "C>T, from allele n2430"
FT mutation
FT 6535
FT /*tag= af
FT /note= "G>A, from allele n2426"
FT intron
FT 6538..7012
FT /*tag= ag
FT /number= Intron_7
FT repeat_unit
FT 6567..6625
FT /*tag= ah
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit
FT 6905..6965
FT /*tag= ai
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon
FT 7013..7075
FT /*tag= aj
FT /number= Exon_8
FT mutation
FT 7020
FT /*tag= ak
FT /note= "C>T, from allele n1163"
FT W09325685-A.
FT 23-DEC-1993.
FT PD 23-DEC-1993.
FT PF 14-JUN-1993; U05701.
FT PR 12-JUN-1992; US-897788.
FT PR 20-NOV-1992; US-979638.
FT PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT PI Horvitz HR, Shiham S, Yuan J;
FT DR WPI: 94-007542/01.
FT DR P-PSDB: R47466.
FT PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT PS Claim 2; Fig 4; 127pp; English.
FT CC This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
FT CC is identified as the ced-3 transcript and was most abundant in
FT CC embryos, but was also detected in larvae and young adults, suggesting
FT CC that ced-3 is expressed not only in cells undergoing cell death. The
FT CC four largest introns as well as sequences 5' of the start codon
FT CC contain repetitive elements, some of which have been characterised
FT CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
FT CC and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
FT CC highly hydrophilic with no significant hydrophobic region that might
FT CC be a transmembrane region. One region of Ced-3 is very rich in serine.
FT CC It is thought that this region is involved in protein-protein inter-
FT CC actions, similar to acid blobs in transcription factors. Of the
FT CC mutations which occur within the ced-3 gene, eight of the mutations
FT CC are missense mutations, two are nonsense mutations and two are putative

```



```

FT FT /*tag= p
FT FT /rpt_type= INVERTED
FT FT /note= Repeat 3"
FT FT 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT FT 5757
FT FT /*tag= r
FT FT /note= "G>A, from allele n2433"
FT FT 5761..5814
FT FT /*tag= s
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= t
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= u
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= v
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /*tag= w
FT FT /number= Exon_7
FT FT 6538..7012
FT FT /*tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= z
FT FT /rpt_type= INVERTED
FT FT /note= Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= aa
FT FT /number= Exon_8
FT FT W0325685-A.
FT PN 23-DEC-1993.
FT PD 14-JUN-1993; U05701.
FT PF 12-JUN-1992; US-897788.
FT PR 20-NOV-1992; US-979638.
FT PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT PI Horvitz HR, Shaham S, Yuan J;
FT DR WPI: 94-007542/01.
FT DR P-PSDB: R53281.
FT PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT PS The sequences given in 064735-45 represent mutations of the C. elegans
FT CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT CC was most abundant in embryos, but was also detected in larvae and young
FT CC adults, suggesting that ced-3 is expressed not only in cells undergoing
FT CC cell death. The four largest introns as well as sequences 5' of the
FT CC start codon contain repetitive elements, some of which have been
FT CC characterised in non-coding regions of other C. elegans genes, such
FT CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT CC region that might be a transmembrane region. One region of Ced-3 is
FT CC very rich in serine. It is thought that this region is involved in
FT CC protein-protein interactions, similar to acid blobs in transcription
FT CC factors. Of the mutations which occur within the ced-3 gene, eight of
FT CC the mutations are missense mutations, two are nonsense mutations and
FT CC two are putative splicing mutations. These mutations establish the
FT CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT CC function is not essential for viability. The ced-3 and ced-4 gene
FT CC products may be used to develop agents for treating conditions
FT CC characterised by cell deaths, such as myocardial infarction, stroke,
FT CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT CC infection, aging or hair loss.
FT CC Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

```

```

Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 901 gtggtcttggtatcctcaacttgccgggtttgtttgttggtacactctccgtgatgc 960
QY |||||
QY 901 GTCGTCTTGGTATPCCTCAACTTGTCCGGTGTGTTTGGTGGTACACTCTCTCCGTGATGC 960
Db 961 caccgtctccgtctcaattatcggttagaatagtgaactgtccagatgggtgaactcata 1020
QY |||||
QY 961 CACCTGTCTCCGTCTCAATTATCGTTTAGAAATGTGAACGTGCAGATGGGTGACTCAATA 1020
Db 1021 ttgctgctgctcaatccactttctttcttcacgcgcagctcttaccagagcccatcaaac 1080
QY |||||
QY 1021 TTGCTGCTGCTACAATCCACTTCTTCTTCATCGGCAGCTTCATACGAGCCCATCAATAAC 1080
Db 1081 tttttttccggcaaaatttgcaataaaacggcgcaaaattttctccaaattgttagcga 1140
QY |||||
QY 1081 TTTTTCGCGGAAATTTGCAATAAACCGCGCAAAACTTCTCCAAATTGTGTAGCGCAA 1140
Db 1141 tatatacaatccataagaatatcttctcaatgtttatgattcttcgcagacattctct 1200
QY |||||
QY 1141 TATATACAATCCATAAGAATATCTCTCAATGTTTATGATTTCTTCGCGAGCACTTCTCT 1200
Db 1201 tcgtgctgaacactcttattttataataatttcgcgtaaaattccgatttttgagtatta 1260
QY |||||
QY 1201 TCGTGTGCTAACATCTTATTTTATAAATTTCCGCTAAATTCGATTTTGTGAGTATTA 1260
Db 1261 atttaccgtaaaattatgataagcgcgaacacacaaactactataaaatggtaaaagctcc 1320
QY |||||
QY 1261 ATTTATCGTAAATTAATATCAATATAGCAGCGAAACTACTTAAATAATGGTAAAGCTCC 1320
Db 1321 taaatcggtcgcacattatcgatttaaggaaatcacaaattctgagaaatgcgactgcgc 1380
QY |||||
QY 1321 TAAATCGGCTCGACATTAATCGTATTAAAGGAATCACAAATTCAGAGATCGTACTGGCG 1380
Db 1381 aacataattgacggcaaaatatctcgtagcgaaaaactacagtaattctttaaagtactac 1440
QY |||||
QY 1381 AACATATTTGACGGCAAAATATCTCGTAGCGAAACTACAGTAATTCCTTAAATGACTAC 1440
Db 1441 tgtagcgttggtgctgattacgggctcaatttttgaaataaatttttttttgcgaattt 1500
QY |||||
QY 1441 TGTAGCGCTTGTGTCGATTACGGGCTCAATTTTGAATAAATTTTTCGAAATTT 1500
Db 1501 tgaatacccgtaaatcgtcaacacgctacagtagctcatttaaaggattactgtagtctta 1560
QY |||||
QY 1501 TGATAACCGGTAAATCGTCACAACGCTACAGTAGTCAATTTAAAGGATTACTGTAGTTCTA 1560
Db 1561 gctacgacataatttgcgcgcaaaatatgactgtaataacgcattctctgaattttgtgtt 1620
QY |||||
QY 1561 GCTACGAGATATTTTTCGCGCGCAAAATATGACTGTAATACGCAATTCCTGTAATTTGTGTT 1620
Db 1621 tccgtaataatttcacaagattttggcattccacttttaaggcgacagagattattcca 1680
QY |||||
QY 1621 TCCGTAATAATTTTCACAGATTTTGGCAATTCACCTTTAAAGCGCACAGGATTTATTCCA 1680
Db 1681 atgggtctggcgcaaggaaaaagttgtatagacttttaaatctccttcgacttttaattc 1740
QY |||||
QY 1681 ATGGGTCTCGGCACGCAAAAAGTTGTATAGACTTTTAAATTCCTCTTGCAATTTTAATTC 1740
Db 1741 aattactaaaattttcggaatttttctgttaaaatttttaaaatcagttttctaaatt 1800
QY |||||
QY 1741 AATTACTAAAATTTTCGTGAATTTTCTGTGTAATTTTAAATAATTTTAAATAATTCGATTT 1800
Db 1801 ttccaggctgacaaacagaaaaacacacacacacacacacacacacacacacacacacac 1860
QY |||||
QY 1801 TTCCAGGCTGACAAACAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1860
Db 1861 taaaaataaacgatttctcattgaaattgtgtttattgtt 1900
QY |||||
QY 1861 TAAAAATAACGATTTCTCATTGAAAAATTTGTGTTTATGTT 1900

```

```

RESULT 14
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DE 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /tag= a
FT FT /label= Repeat-1
FT FT 1490..1614
FT FT /tag= b
FT FT /label= Repeat-2
FT FT 2167..2366
FT FT /tag= c
FT FT /codon_start= 2232..2234
FT FT 2310
FT FT /tag= d
FT FT /label= T(n1040)
FT FT /note= "causes L27F mutation"
FT FT 2367..2429
FT FT /tag= e
FT FT /label= Intron-1
FT FT 2430..2575
FT FT /tag= f
FT FT 2487
FT FT /tag= g
FT FT /label= A(n718)
FT FT /note= "causes G65R mutation"
FT FT 2576..2853
FT FT /tag= h
FT FT /label= Intron-2
FT FT 2854..3107
FT FT /tag= i
FT FT 3108..4302
FT FT /tag= j
FT FT /label= Intron-3
FT FT 3126..3243
FT FT /tag= k
FT FT /label= Repeat-1
FT FT 3329..3396
FT FT /tag= l
FT FT /label= Repeat-1
FT FT 3487..3759
FT FT /tag= m
FT FT /label= Repeat-2
FT FT 3782..4070
FT FT /tag= n
FT FT /label= Repeat-2
FT FT 4303..4633
FT FT /tag= o
FT FT 4634..5546
FT FT /tag= p
FT FT /label= Intron-4
FT FT 4688..4719
FT FT /tag= q
FT FT /label= Repeat-3
FT FT 5221..5330
FT FT /tag= r
FT FT /label= Repeat-3
FT FT 5546..5760
FT FT /tag= s
FT FT 5757
FT FT /tag= t
FT FT /label= A(n2433)
FT FT /note= "causes G360S mutation"
FT FT 5761..5814
FT FT /tag= u
FT FT /label= Intron-5
FT FT 5815..5942

```

```

FT allele /tag= v
FT 5940
FT /tag= w
FT /label= T(n1165)
FT /note= "Creates premature stop codon at 403"
FT 5943..6297
FT /tag= x
FT /label= Intron-6
FT 6062..6138
FT /tag= y
FT /label= Repeat-4
FT 6298..6537
FT /tag= z
FT 6322
FT /tag= aa
FT /label= T(n1949)
FT /note= "creates premature stop codon at 412"
FT 6372
FT /tag= ab
FT /label= A(n1286)
FT 6434
FT /tag= ac
FT /label= T(n1129,n1164)
FT /note= "causes A449V mutation"
FT 6485
FT /tag= ad
FT /label= T(n2430)
FT /note= "causes A466V mutation"
FT 6535
FT /tag= ae
FT /label= A(n2426)
FT /note= "causes E483K mutation"
FT 6538..7012
FT /tag= af
FT /label= Intron-7
FT 6567..6625
FT /tag= ag
FT /label= Repeat-5
FT 6905..6965
FT /tag= ah
FT /label= Repeat-5
FT 7013..7653
FT /tag= ai
FT /note= "in-frame stop codon at 7073-7075"
FT 7020
FT /tag= aj
FT /label= T(n1163)
FT /note= "causes S486F mutation"
FT W09625946-A1.
FT 29-AUG-1996.
FT 23-FEB-1996; U02473.
FT 24-FEB-1996; US-394189.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI; 96-425082/42.
FT P-PSDB; R38754.
FT Ced-3 and human interleukin 1-beta convertase genes and proteins
FT useful to treat inflammation and diseases characterised by cell
FT death
FT Claim 18; Fig 3; 139pp; English.
FT The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
FT mapping RFLPs and chromosome walking, and genomic DNA cloned
FT in plasmid pJ107 was sequenced. EMS-induced alleles were also
FT sequenced. The gene codes for a cell death protein (R38754) that
FT is structurally similar to human interleukin-1 beta converting
FT enzyme (ICE) (R38755), suggesting that ced-3 protein may be a
FT cysteine protease like ICE and that ICE may be a human equivalent
FT of the nematode cell death gene. The ced-3 gene can be used as a
FT probe or in the prodn. of Ced-3 protein and novel drugs for
FT enhancing or inhibiting the activity of ICE, ced-3 and related
FT genes for the treatment of inflammatory diseases and/or diseases
FT caused by cell death. Novel inhibitors of ced-3 activity include
FT portions of the ced-3 gene and it product.

```

```
SQ Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;
Query Match 14.5%; Score 145; DB 24; Length 7653;
Best Local Similarity 83.3%; Pred. No. 1.27e-62;
Matches 244; Conservative 0; Mismatches 39; Indels 10; Gaps 8;

Db 1342 tattaaggaatcacaaatctgagaatgagctgctgagcaacatatttgacg-gcaaaat 1400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1629 TATTACGGAAACAAATATCAGAGATGCGTATTCAGT-CATATTGGCGCGCAAAAT 1571
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1401 atctcgtaggaatacactagtaattcttaaatgactactgtagcgcttgctgattt 1460
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1570 ATCTCGTAGCTAGACTACAGTATCTTTAAATGACTACTGTAGCG-TTGTGACGATTT 1512
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1461 acgggct--caattttgaaaaataatttttttttogaattttgataaccgcgtaaatcgt 1518
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1511 ACGGGTTATCAAAATTCGAAAAAATAATTTTCAAAAATTCAG--CCCGTAAATCGA 1454
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1519 cacaac-gcctacagtagctatttaagagattactgtagttctagctacgagatatttgc 1577
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1453 CACAAGCGCTACAGTAGTCATTTAAGAATTAAGTACTGTAGTTTTTCGCTACGAGATATTTGC 1394
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1578 gcgcgaatatgact-gtaatacgcattctctgaattttgttttccgtaata 1629
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1393 -CGTCAAAATATGTTGCGCAGTACGATCTTCAGAAATTTGTGATTCTTAAATA 1342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
ID Q64744 standard; DNA; 7653 BP.
AC Q64744;
DT 23-JUN-1994 (first entry)
UN ced-3 (G6536A) gene.
KW C. elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /*tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /*tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /*tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /*tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /*tag= g
FT /number= Exon_3
FT intron 3108..4302
FT /*tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
```

```
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /*tag= m
FT /number= Exon_4
FT intron 4635..5546
FT /*tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /*tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /*tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /*tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /*tag= t
FT /number= Intron_6
FT repeat_region 6062..6138
FT /*tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /*tag= v
FT /number= Exon_7
FT mutation 6538
FT /*tag= w
FT /note= "G>A, from allele n2426"
FT intron 6538..7012
FT /*tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /*tag= aa
FT /number= Exon_8
FT WO9325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI; 94-007542/01.
FT P-PSDB; R53287.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; fig 4; 127pp; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT
```

WATERMAN

(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MFsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 02:01:37 1999; WasPar time 1697.90 Seconds
Tabular output not generated. 1380.034 Million cell updates/sec

Title: >US-08-287-669-18
Description: (901-1900) from US08287669.seq (2 of 10)
Perfect Score: 1000
N.A. Sequence: 901 GTCGTCCTGGTATCCTCAA.....TGAAAAATGTTTATGTT 1900
Comp: CAGCAGGAACCATAGGAGTT.....ACTTTTACACAAAATACAA

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 2883791 seqs, 1171580779 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est11
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33
33:gb_est34 34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38
38:gb_est39 39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.523; Variance 3.278; scale 3.515
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES																	
Result	No.	Score	%	Query	Match	Length	DB	ID	Description	Pred. No.							
C	1	91	9.1	359	8	T01235	WEST01956	Early embryo	5.48e-81								
	2	82	8.2	359	8	T01235	WEST01956	Early embryo	4.99e-69								
C	3	68	6.8	300	15	C52036	C52036	Yuji Kohara unp	7.47e-51								
C	4	55	5.5	360	32	DY3308	CELSK098H4F	Yuji Kohara	1.26e-34								
C	5	54	5.4	252	17	AA754459	97SN1787	Rice Immature	2.06e-33								
C	6	54	5.4	357	15	C41501	C41501	Yuji Kohara unp	2.06e-33								
C	7	51	5.1	337	32	D64660	CELSK086D2R	Yuji Kohara	8.30e-30								
C	8	50	5.0	252	17	AA754459	97SN1787	Rice Immature	1.28e-28								
C	9	50	5.0	300	15	C29898	C29898	Yuji Kohara unp	1.28e-28								
10	49	4.9	357	15	C41501	C41501	Yuji Kohara unp	1.95e-27									

C	11	46	4.6	247	17	AA754458	97SN1784 Rice	Immature	6.24e-24
C	12	46	4.6	300	15	C32017	C32017 YUJI Kohara unp	6.24e-24	
C	13	43	4.3	300	15	C53842	CELK098H4F YUJI Kohara	1.69e-20	
C	14	43	4.3	360	32	D75308	CELK098H4F YUJI Kohara	1.69e-20	
C	15	42	4.2	300	15	C52036	C52036 YUJI Kohara unp	2.26e-19	
C	16	41	4.1	247	17	AA754458	97SN1784 Rice	Immature	2.96e-18
C	17	39	3.9	337	32	D64660	CELK086D2R YUJI Kohara	4.73e-16	
C	18	38	3.8	300	15	C29898	C29898 YUJI Kohara unp	5.75e-15	
C	19	36	3.6	300	15	C32017	C32017 YUJI Kohara unp	7.86e-13	
C	20	35	3.5	300	15	C53842	C53842 YUJI Kohara unp	8.80e-12	
C	21	34	3.4	300	15	C29734	C29734 YUJI Kohara unp	9.54e-11	
C	22	34	3.4	2275	20	AF034173	AF034173 Human mRNA (T	9.54e-11	
C	23	33	3.3	306	22	AI082072	os2g03.xl Soares_sene	1.00e-09	
C	24	33	3.3	356	23	AI183776	qei7h08.xl Soares_feta	1.00e-09	
C	25	33	3.3	374	20	AA918212	on82g01.s1 Soares_NFL	1.00e-09	
C	26	33	3.3	400	8	M79744	WBT00281 Mixed stage,	1.00e-09	
C	27	33	3.3	424	16	AA622274	no42h08.s1 NCI_CGAP.Pr	1.00e-09	
C	28	33	3.3	525	24	AI188773	q416f05.xl Soares_plac	1.00e-09	
C	29	32	3.2	300	15	C35709	C35709 YUJI Kohara unp	1.01e-08	
C	30	31	3.1	215	26	AA381412	tc50f01.xl Soares_tota	9.88e-08	
C	31	31	3.1	225	18	AA766012	oal1c02.s1 NCI_CGAP_GC	9.88e-08	
C	32	31	3.1	244	24	AI198459	qf88a07.xl Soares_plac	9.88e-08	
C	33	31	3.1	315	21	AI016448	ot78f05.s1 Soares_tota	9.88e-08	
C	34	31	3.1	326	29	AI568828	th16f10.xl NCI_CGAP.Pr	9.88e-08	
C	35	31	3.1	417	23	AI149055	qc81b08.xl Soares_plac	9.88e-08	
C	36	31	3.1	429	25	AI299057	qn35e09.xl NCI_CGAP_KI	9.88e-08	
C	37	31	3.1	429	23	AI093094	qg97a10.xl Soares_feta	9.88e-08	
C	38	31	3.1	429	23	AI151518	qc80b12.xl Soares_plac	9.88e-08	
C	39	31	3.1	440	20	AA907140	od44c05.s1 NCI_CGAP_GC	9.88e-08	
C	40	31	3.1	441	23	AI138427	qg84d07.xl Soares_test	9.88e-08	
C	41	31	3.1	477	18	AA804752	ob98e04.s1 NCI_CGAP_GC	9.88e-08	
C	42	31	3.1	494	35	AA039776	zf10c10.s1 Soares_feta	9.88e-08	
C	43	31	3.1	505	17	AA722943	zg85a06.s1 Soares_feta	9.88e-08	
C	44	31	3.1	2275	20	AF034173	AF034173 Human mRNA (T	9.88e-08	
C	45	30	3.0	563	23	AI146823	qb92f01.xl Soares_feta	9.26e-07	

ALIGNMENTS

RESULT 1
LOCUS T01235 359 bp mRNA 10-NOV-1992
DEFINITION WEST01956 Early embryo, StrataGene (cat. #937007) Caenorhabditis elegans CDNA clone CEES041, mRNA sequence.
ACCESSION T01235
NID 9277716
VERSION T01235.1 GI:277716
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 359)
AUTHORS McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M., Fitzgerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R., Venter,J.C. and Fields,C.A.
TITLE Caenorhabditis elegans cDNAs
JOURNAL Unpublished (1993)
COMMENT Other_ESTs: WEST01957
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetigr.org
Seq primer: M13 Forward
Location/Qualifiers
1. .359
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="CEES041"
/clone_lib="Early embryo, StrataGene (cat. #937007)"
BASE COUNT 115 a 61 c 60 g 122 t 1 others

ORIGIN

Query Match 9.1%; Score 91; DB 8; Length 359;
 Best Local Similarity 86.3%; Pred. No. 5.48e-81;
 Matches 113; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

Db 1 CCGGTAATCGACACAGCCNTACAGTACTGATTTAAAGGATTACTGTAGTTTCGCTAT 60
 |||||
 Cp 1465 CCGGTAATCGACACAGCCGCTACAGTACTGATTTAAAGGATTACTGTAGTTTCGCTAT 1406
 |||||

Db 61 GAGATATTTTCGCGTGAATATGGTGTGCAATACGATATTCCTGAAATTTTCGGTTTCAG 120
 |||||
 Cp 1405 GAGATATTTTCGCGTGAATATGGTGTGCAATACGATATTCCTGAAATTTTCGGTTTCAG 1347
 |||||

Db 121 TAATACATAAA 131
 |||||

Cp 1346 TAATACGATAA 1336

RESULT 2
 LOCUS T01235 359 bp mRNA EST 10-NOV-1992
 DEFINITION WEST01956 Early embryo, Stragatene (cat. #937007) Caenorhabditis
 elegans cDNA clone CEES041, mRNA sequence.

ACCESSION T01235
 NID 9277716
 VERSION T01235.1 GI:277716
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 359)
 McCombie, W.R., Kelley, J.M., Aubin, L., Goscoechea, M.,
 Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R.,
 Venter, J.C. and Fields, C.A.
 Caenorhabditis elegans cDNAs
 Unpublished (1993)
 Other ESTs: WEST01957
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@tigr.org
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..359
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="CEES041"

BASE COUNT 115 a 61 c 60 g 122 t 1 others
 ORIGIN

Query Match 8.2%; Score 82; DB 8; Length 359;
 Best Local Similarity 86.4%; Pred. No. 4.99e-69;
 Matches 108; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

Db 1 CCGGTAATCGACACAGCCNTACAGTACTGATTTAAAGGATTACTGTAGTTTCGCTAT 60
 |||||
 Qy 1507 CCGGTAATCGTCACAA-CGCTACAGTAGTCAATTTAAAGGATTACTGTAGTTTCGCTAC 1565
 |||||

Db 61 GAGATATTTTCGCGTGAATATGGTGTGCAATACGATATTCCTGAAATTTTCGGTTTCAG 120
 |||||

Qy 1566 GAGATATTTTCGCGCAAAATATGAC-TGTAATACGATTCCTGAAATTTTCGGTTTCG 1624
 |||||

Db 121 TAATA 125
 |||||

Qy 1625 TAATA 1629
 |||||

RESULT 3

LOCUS C52036 300 bp mRNA EST 11-SEP-1997
 DEFINITION C52036 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk2259l 3', mRNA sequence.

ACCESSION C52036
 NID 92389793
 VERSION C52036.1 GI:2389793
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 300)
 Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
 Sano, M., Miyata, A. and Nishigaki, A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 On Nov 29, 1993 this sequence version replaced gi:430548.

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers
 1..300
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /note="dev_stage=varied, sex=Hermaphrodite male,
 tissue_type=whole animal"
 /db_xref="taxon:6239"
 /map="21"
 /clone="yk2259l"
 /clone_lib="Yuji Kohara unpublished cDNA"
 115 a 44 c 52 g 87 t 2 others

FEATURES
source

Query Match 6.8%; Score 68; DB 15; Length 300;
 Best Local Similarity 91.5%; Pred. No. 7.47e-51;
 Matches 75; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 178 TTTTCAAAAATCGAGCCGTTAAATCGACACAGCTCTACAGTAAGTAATTAAGAATTAC 237
 |||||

Cp 1480 TTTTCAAAAATCGAGCCGTTAAATCGACACAGCGTACAGTAGTCATTTTAAAGAATTAC 1421
 |||||

Db 238 TGTAGTTTTTGTCTACGAGATAT 259
 |||||

Cp 1420 TGTAGTTTTTGTCTACGAGATAT 1399
 |||||

RESULT 4
 LOCUS D75308 360 bp mRNA EST 14-DEC-1995
 DEFINITION CELK098H4F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk98h4 5', mRNA sequence.

ACCESSION D75308
 NID g1121092
 VERSION D75308.1 GI:1121092
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 360)
 Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
 Tabara, H.
 Toward an expression map of the C.elegans genome
 Unpublished (1994)

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan

Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp

FEATURES

source
1..300
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk228a6"
/clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT 111 a 42 c 60 g 86 t 1 others
ORIGIN

Query Match 5.0%; Score 50; DB 15; Length 300;
Best Local Similarity 86.6%; Pred. No. 1.28e-28;
Matches 71; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

Db 97 GAAATCGCTATTGTCAGCATATTTCACGAGCAAAATATTGTCGCGAAACTACAG-155

Qy 1364 GAGAAATCGCTACTGCGCAACATATTGACG-GCAAAATATCTCTGACGAAACTACAGT 1422

Db 156 AATTCCTCAATGACTACTGTA 177

Qy 1423 AATTCCTTAATGACTACTGTA 1444

RESULT 10
LOCUS C41501 357 bp mRNA EST 10-SEP-1997
DEFINITION C41501 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk268f11 5', mRNA sequence.

ACCESSION C41501
NID 92377738
VERSION C41501.1 GI:23777738
KEYWORDS EST.
SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 357)
AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402115.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp

FEATURES
source
1..357
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk268f11"
/clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT 95 a 90 c 49 g 123 t
ORIGIN

Query Match 4.9%; Score 49; DB 15; Length 357;
Best Local Similarity 86.5%; Pred. No. 1.95e-27;
Matches 64; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Db 2 CCCGTGATCGACATCATCGCTACAGTAGTCATTAAAGGGTTACTGTAGTTTCCTGCTG 61
Qy 1507 CCCGTAATCGTCA-CAACGCTAGTAGTCATTAAAGGATTACTGTAGTTCTAGCTAC 1565
Db 62 GAGATATTGGCGC 75
Qy 1566 GAGATATTGGCGC 1579

RESULT 11
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
cDNA clone 97SN1784, mRNA sequence.

ACCESSION AA754458
NID 92801164
VERSION AA754458.1 GI:2801164
KEYWORDS EST.
SOURCE Oryza sativa.

ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 247)
AUTHORS Nam, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P.,
Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y.,
Lee, M.C. and Eun, M.Y.

TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nam, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.

FEATURES

source
1..247
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 7 a 16 c 21 g 34 t 169 others
ORIGIN

Query Match 4.6%; Score 46; DB 17; Length 247;
Best Local Similarity 12.2%; Pred. No. 6.24e-24;
Matches 27; Conservative 102; Mismatches 90; Indels 2; Gaps 2;

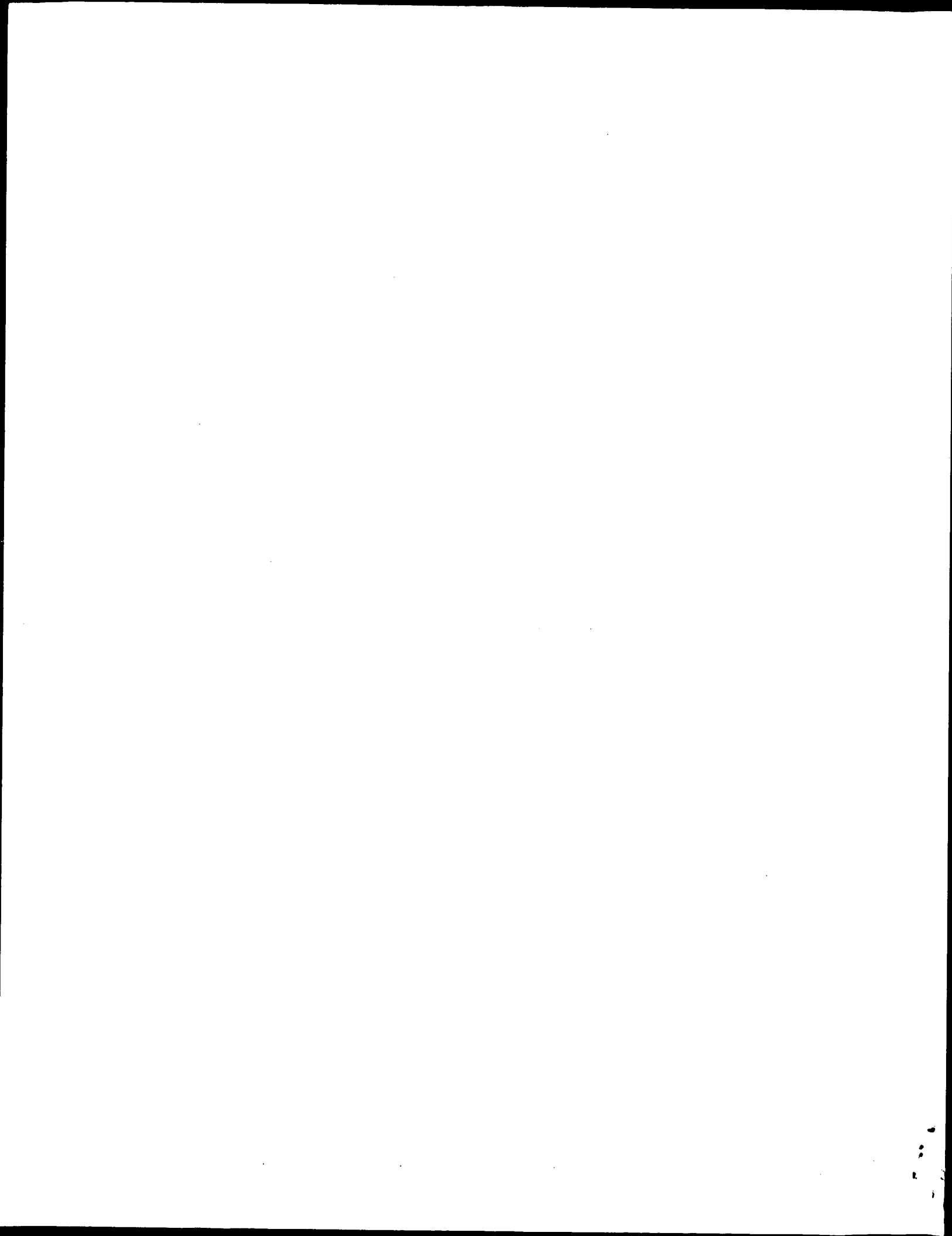
Db 20 MNKHTHMTBBWCVRVGTNNKHNGRTTWNDCSDNAHCRVYVBYVYARSKYGT 79
Cp 1137 CGTAACAATTGGAGAAAGTTTTCGCGGTTTATTGCAAAATTCGCGAAAAAAGTT 1078

Db 80 BYYSWNVDTNIGTGKGTNNVHSGWNNRCSNVVYVWBTAYCDY-BHYBDRANHVDT 138
Cp 1077 TATGATGGCTCGTAAGACTGCGCATGAGAAAAGTGGATTGTAGCAGCAGCAATAT 1018

Db 139 RCINDRGYCNVTASDNGTSATKRVTDKTDSDCGGCGKRVKVTYSSBYBRGVNVVRT 198
Cp 1017 GAGTCACCCATCTGGACAGAGTTCACATTTCTAAACGATAATTGAGACGGAGAC-AGGTGGC 959

ASE COUNT	107 a	81 c	52 g	119 t	1 others
RIGIN					

Search completed: Fri Aug 6 02:59:59 1999
Job time : 3502 secs.



WQSEFH
(TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Aug 7 20:29:46 1999; MasPar time 1829.25 Seconds
Tabular output not generated.

Title: >US-08-287-669-18
Description: (5401-6400) from US08287669.seq (7 of 10)
Perfect Score: 1000
N.A. Sequence: 5401 ATTCAAAAAAGTGAAT.....TGCTCGTGATCATGTTCA 6400
Comp: TAAGTTTTTTTCAGCTTA.....ACGAGCACCTAGTACCAAGT

Scoring table:
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
Database: genbank11
17:gb_ba1 18:gb_ba2 19:gb_bt1 20:gb_bt2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 11.316; Variance 7.880; scale 1.436
pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Description	Pred. No.	
1	1000	100.0	CELCE3A	Caenorhabditis elegans	0.00e+00
2	957	95.7	CELCE3A	Caenorhabditis elegans	0.00e+00
3	955	95.5	CELCE3A	Caenorhabditis elegans	0.00e+00
4	75	7.5	CELCE3A	Caenorhabditis elegans	9.31e-23
5	75	7.5	CELCE3A	Caenorhabditis elegans	9.51e-23
6	75	7.5	CELCE3A	Caenorhabditis elegans	9.51e-23
7	75	7.5	CELCE3A	Caenorhabditis elegans	9.51e-23
8	74	7.4	CELCE3A	Caenorhabditis elegans	3.27e-22
9	74	7.4	CELCE3A	Caenorhabditis elegans	3.27e-22
10	74	7.4	CELCE3A	Caenorhabditis elegans	3.27e-22
11	72	7.2	CELCE3A	Caenorhabditis elegans	3.82e-21
12	72	7.2	CELCE3A	Caenorhabditis elegans	3.82e-21
13	69	6.9	CELCE3A	Caenorhabditis elegans	1.49e-19

14	69	6.9	115355	19	CEY55D9	Caenorhabditis elegans	1.49e-19
15	69	6.9	224746	19	CEY56A3	Caenorhabditis elegans	1.49e-19
16	68	6.8	23961	22	CELR02C2	Caenorhabditis elegans	5.04e-19
17	68	6.8	38876	22	CELR119	Caenorhabditis elegans	5.04e-19
18	68	6.8	91286	21	CEY102A5C	Caenorhabditis elegans	5.04e-19
19	68	6.8	299202	19	AC006792	Caenorhabditis elegans	5.04e-19
20	68	6.8	299782	19	AC006844	Caenorhabditis elegans	5.04e-19
21	68	6.8	312267	19	AC006785	Caenorhabditis elegans	5.04e-19
22	67	6.7	38103	21	CELR09A1	Caenorhabditis elegans	1.69e-18
23	67	6.7	108589	19	AC006795	Caenorhabditis elegans	1.69e-18
24	67	6.7	132742	21	CEY37D8A	Caenorhabditis elegans	1.69e-18
25	67	6.7	150641	21	CEY41E3	Caenorhabditis elegans	1.69e-18
26	67	6.7	257866	19	AC006794	Caenorhabditis elegans	1.69e-18
27	66	6.6	19897	19	AC006623	Caenorhabditis elegans	5.68e-18
28	66	6.6	26195	21	CEY0466	Caenorhabditis elegans	5.68e-18
29	66	6.6	40662	21	CEY32B12B	Caenorhabditis elegans	5.68e-18
30	66	6.6	42799	22	CELF56F11	Caenorhabditis elegans	5.68e-18
31	66	6.6	125590	21	CEY49E10	Caenorhabditis elegans	5.88e-18
32	66	6.6	191857	19	CEY32B12	Caenorhabditis elegans	5.88e-18
33	66	6.6	244239	19	AC006880	Caenorhabditis elegans	5.88e-18
34	66	6.6	298406	21	CEY75B8A	Caenorhabditis elegans	5.88e-18
35	66	6.6	314495	19	CEY76H2	Caenorhabditis elegans	5.88e-18
36	66	6.6	336638	19	CEY75B8	Caenorhabditis elegans	5.88e-18
37	65	6.5	42799	22	CELF56F11	Caenorhabditis elegans	1.90e-17
38	65	6.5	48088	22	CELY44B3A	Caenorhabditis elegans	1.90e-17
39	65	6.5	244239	19	AC006880	Caenorhabditis elegans	1.90e-17
40	65	6.5	298960	20	AC006896	Caenorhabditis elegans	1.90e-17
41	65	6.5	319857	19	AC006782	Caenorhabditis elegans	6.32e-17
42	64	6.4	70594	19	AC006708	Caenorhabditis elegans	6.32e-17
43	64	6.4	110000	19	CEY111B2	Caenorhabditis elegans	6.32e-17
44	64	6.4	299670	19	AC006845	Caenorhabditis elegans	6.32e-17
45	63	6.3	224525	19	CEY70C5	Caenorhabditis elegans	2.09e-16

ALIGNMENTS

RESULT	1	CELCE3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS		Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
ACCESSION		L29052				
NID		9456416				
VERSION		L29052.1	GI:456416			
KEYWORDS		cell death protein; interleukin-1 beta converting enzyme.				
SOURCE		Caenorhabditis elegans (strain N2) DNA.				
ORGANISM		Caenorhabditis elegans				
REFERENCE		Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.				
AUTHORS		Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.				
TITLE		The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1b-converting enzyme				
JOURNAL		Cell 75, 641-652 (1993)				
MEDLINE		94061982				
FEATURES		Location/Qualifiers				
Source		1. 7653				
		/organism="Caenorhabditis elegans"				
		/strain="N2"				
		/db_xref="taxon:6239"				
		2167. 2366				
		/gene="ced-3"				
		/number=1				
		join(2232. 2366,2430. 2575,2854. 3107,4303. 4634, 5547. 5760,5815. 5942,6298. 6537,7013. 7075)				
		/gene="ced-3"				
		join(2232. 2366,2430. 2575,2854. 3107,4303. 4634, 5547. 5760,5815. 5942,6298. 6537,7013. 7075)				
		/gene="ced-3"				
		/note="codes for a protein similar to mammalian interleukin-1b-converting enzyme"				
		/codon_start=1				
		/product="cell death protein"				

```

/protein_id="AA27982.1"
/db_xref="PID:9456417"
/db_xref="GI:456417"
/translaton="MMRODRSLRNIMFSSHLKVDLEVLIAKOVLSNDGDMT
NSCGTVREKREIVKAVORRQDVFADAFYDALRSTGHEGLAEVLEPLARSVDNSAVEF
ECMSPASHRRSRALPAGVTSPTVRHRSVSVSASF SYQDIYSRARSRSRALHS
SDRNTYSSPPVNAFPSPQSSANSFTGCSLSYSSRNRSFSKASGPTQYIFBEDMN
FVDAPISRVEDKTYRNFSSPRGMCLINNEHEFOEMPTNRNGTKADKDNLTNLFROM
GYTVICKDNLTKGMULTIRDFAKHSHGDSAILVLSHGEENVLIIGVDDIPISTHEI
YDLNANAPRLANKPKIYVQACRGERDNGFPVLDSDVGPAPFLRGWDRGDPLE
NFGCVRPQVQWRKPKSQADILIRYATTAQYVSWRNSARGSMFIOACVEVSTHAK
DMDVVELLTVKVKVACGFGTSGSNNLKQMFEMTSRLLLKKFYFWPERNASV"
intron
2367..2429
/feature="ced-3"
/number=1
2430..2575
/feature="ced-3"
/number=2
2576..2853
/feature="ced-3"
/number=2
2854..3107
/feature="ced-3"
/number=3
3108..4302
/feature="ced-3"
/number=3
4303..4634
/feature="ced-3"
/number=4
4635..5546
/feature="ced-3"
/number=4
5547..5760
/feature="ced-3"
/number=5
5761..5814
/feature="ced-3"
/number=5
5815..5942
/feature="ced-3"
/number=6
5943..6297
/feature="ced-3"
/number=6
6298..6537
/feature="ced-3"
/number=7
6538..7012
/feature="ced-3"
/number=7
7013..7652
/feature="ced-3"
/number=8
7653..8129
/feature="ced-3"
/number=8
8130..8484
/feature="ced-3"
/number=9
8485..8840
/feature="ced-3"
/number=9
8841..9196
/feature="ced-3"
/number=10
9197..9552
/feature="ced-3"
/number=10
9553..9908
/feature="ced-3"
/number=11
9909..10264
/feature="ced-3"
/number=11
10265..10620
/feature="ced-3"
/number=12
10621..10976
/feature="ced-3"
/number=12
10977..11332
/feature="ced-3"
/number=13
11333..11688
/feature="ced-3"
/number=13
11689..12044
/feature="ced-3"
/number=14
12045..12400
/feature="ced-3"
/number=14
12401..12756
/feature="ced-3"
/number=15
12757..13112
/feature="ced-3"
/number=15
13113..13468
/feature="ced-3"
/number=16
13469..13824
/feature="ced-3"
/number=16
13825..14180
/feature="ced-3"
/number=17
14181..14536
/feature="ced-3"
/number=17
14537..14892
/feature="ced-3"
/number=18
14893..15248
/feature="ced-3"
/number=18
15249..15604
/feature="ced-3"
/number=19
15605..15960
/feature="ced-3"
/number=19
15961..16316
/feature="ced-3"
/number=20
16317..16672
/feature="ced-3"
/number=20
16673..17028
/feature="ced-3"
/number=21
17029..17384
/feature="ced-3"
/number=21
17385..17740
/feature="ced-3"
/number=22
17741..18096
/feature="ced-3"
/number=22
18097..18452
/feature="ced-3"
/number=23
18453..18808
/feature="ced-3"
/number=23
18809..19164
/feature="ced-3"
/number=24
19165..19520
/feature="ced-3"
/number=24
19521..19876
/feature="ced-3"
/number=25
19877..20232
/feature="ced-3"
/number=25
20233..20588
/feature="ced-3"
/number=26
20589..20944
/feature="ced-3"
/number=26
20945..21300
/feature="ced-3"
/number=27
21301..21656
/feature="ced-3"
/number=27
21657..22012
/feature="ced-3"
/number=28
22013..22368
/feature="ced-3"
/number=28
22369..22724
/feature="ced-3"
/number=29
22725..23080
/feature="ced-3"
/number=29
23081..23436
/feature="ced-3"
/number=30
23437..23792
/feature="ced-3"
/number=30
23793..24148
/feature="ced-3"
/number=31
24149..24504
/feature="ced-3"
/number=32
24505..24860
/feature="ced-3"
/number=32
24861..25216
/feature="ced-3"
/number=33
25217..25572
/feature="ced-3"
/number=33
25573..25928
/feature="ced-3"
/number=34
25929..26284
/feature="ced-3"
/number=34
26285..26640
/feature="ced-3"
/number=35
26641..26996
/feature="ced-3"
/number=35
26997..27352
/feature="ced-3"
/number=36
27353..27708
/feature="ced-3"
/number=36
27709..28064
/feature="ced-3"
/number=37
28065..28420
/feature="ced-3"
/number=37
28421..28776
/feature="ced-3"
/number=38
28777..29132
/feature="ced-3"
/number=38
29133..29488
/feature="ced-3"
/number=39
29489..29844
/feature="ced-3"
/number=39
29845..30200
/feature="ced-3"
/number=40
30201..30556
/feature="ced-3"
/number=40
30557..30912
/feature="ced-3"
/number=41
30913..31268
/feature="ced-3"
/number=41
31269..31624
/feature="ced-3"
/number=42
31625..31980
/feature="ced-3"
/number=42
31981..32336
/feature="ced-3"
/number=43
32337..32692
/feature="ced-3"
/number=43
32693..33048
/feature="ced-3"
/number=44
33049..33404
/feature="ced-3"
/number=44
33405..33760
/feature="ced-3"
/number=45
33761..34116
/feature="ced-3"
/number=45
34117..34472
/feature="ced-3"
/number=46
34473..34828
/feature="ced-3"
/number=46
34829..35184
/feature="ced-3"
/number=47
35185..35540
/feature="ced-3"
/number=47
35541..35896
/feature="ced-3"
/number=48
35897..36252
/feature="ced-3"
/number=48
36253..36608
/feature="ced-3"
/number=49
36609..36964
/feature="ced-3"
/number=49
36965..37320
/feature="ced-3"
/number=50
37321..37676
/feature="ced-3"
/number=50
37677..38032
/feature="ced-3"
/number=51
38033..38388
/feature="ced-3"
/number=51
38389..38744
/feature="ced-3"
/number=52
38745..39100
/feature="ced-3"
/number=52
39101..39456
/feature="ced-3"
/number=53
39457..39812
/feature="ced-3"
/number=53
39813..40168
/feature="ced-3"
/number=54
40169..40524
/feature="ced-3"
/number=54
40525..40880
/feature="ced-3"
/number=55
40881..41236
/feature="ced-3"
/number=55
41237..41592
/feature="ced-3"
/number=56
41593..41948
/feature="ced-3"
/number=56
41949..42304
/feature="ced-3"
/number=57
42305..42660
/feature="ced-3"
/number=57
42661..43016
/feature="ced-3"
/number=58
43017..43372
/feature="ced-3"
/number=58
43373..43728
/feature="ced-3"
/number=59
43729..44084
/feature="ced-3"
/number=59
44085..44440
/feature="ced-3"
/number=60
44441..44796
/feature="ced-3"
/number=60
44797..45152
/feature="ced-3"
/number=61
45153..45508
/feature="ced-3"
/number=61
45509..45864
/feature="ced-3"
/number=62
45865..46220
/feature="ced-3"
/number=62
46221..46576
/feature="ced-3"
/number=63
46577..46932
/feature="ced-3"
/number=63
46933..47288
/feature="ced-3"
/number=64
47289..47644
/feature="ced-3"
/number=64
47645..48000
/feature="ced-3"
/number=65
48001..48356
/feature="ced-3"
/number=65
48357..48712
/feature="ced-3"
/number=66
48713..49068
/feature="ced-3"
/number=66
49069..49424
/feature="ced-3"
/number=67
49425..49780
/feature="ced-3"
/number=67
49781..50136
/feature="ced-3"
/number=68
50137..50492
/feature="ced-3"
/number=68
50493..50848
/feature="ced-3"
/number=69
50849..51204
/feature="ced-3"
/number=69
51205..51560
/feature="ced-3"
/number=70
51561..51916
/feature="ced-3"
/number=70
51917..52272
/feature="ced-3"
/number=71
52273..52628
/feature="ced-3"
/number=71
52629..52984
/feature="ced-3"
/number=72
52985..53340
/feature="ced-3"
/number=72
53341..53696
/feature="ced-3"
/number=73
53697..54052
/feature="ced-3"
/number=73
54053..54408
/feature="ced-3"
/number=74
54409..54764
/feature="ced-3"
/number=74
54765..55120
/feature="ced-3"
/number=75
55121..55476
/feature="ced-3"
/number=75
55477..55832
/feature="ced-3"
/number=76
55833..56188
/feature="ced-3"
/number=76
56189..56544
/feature="ced-3"
/number=77
56545..56900
/feature="ced-3"
/number=77
56901..57256
/feature="ced-3"
/number=78
57257..57612
/feature="ced-3"
/number=78
57613..57968
/feature="ced-3"
/number=79
57969..58324
/feature="ced-3"
/number=79
58325..58680
/feature="ced-3"
/number=80
58681..59036
/feature="ced-3"
/number=80
59037..59392
/feature="ced-3"
/number=81
59393..59748
/feature="ced-3"
/number=81
59749..60104
/feature="ced-3"
/number=82
60105..60460
/feature="ced-3"
/number=82
60461..60816
/feature="ced-3"
/number=83
60817..61172
/feature="ced-3"
/number=83
61173..61528
/feature="ced-3"
/number=84
61529..61884
/feature="ced-3"
/number=84
61885..62240
/feature="ced-3"
/number=85
62241..62596
/feature="ced-3"
/number=85
62597..62952
/feature="ced-3"
/number=86
62953..63308
/feature="ced-3"
/number=86
63309..63664
/feature="ced-3"
/number=87
63665..64020
/feature="ced-3"
/number=87
64021..64376
/feature="ced-3"
/number=88
64377..64732
/feature="ced-3"
/number=88
64733..65088
/feature="ced-3"
/number=89
65089..65444
/feature="ced-3"
/number=89
65445..65800
/feature="ced-3"
/number=90
65801..66156
/feature="ced-3"
/number=90
66157..66512
/feature="ced-3"
/number=91
66513..66868
/feature="ced-3"
/number=91
66869..67224
/feature="ced-3"
/number=92
67225..67580
/feature="ced-3"
/number=92
67581..67936
/feature="ced-3"
/number=93
67937..68292
/feature="ced-3"
/number=93
68293..68648
/feature="ced-3"
/number=94
68649..69004
/feature="ced-3"
/number=94
69005..69360
/feature="ced-3"
/number=95
69361..69716
/feature="ced-3"
/number=95
69717..70072
/feature="ced-3"
/number=96
70073..70428
/feature="ced-3"
/number=96
70429..70784
/feature="ced-3"
/number=97
70785..71140
/feature="ced-3"
/number=97
71141..71496
/feature="ced-3"
/number=98
71497..71852
/feature="ced-3"
/number=98
71853..72208
/feature="ced-3"
/number=99
72209..72564
/feature="ced-3"
/number=99
72565..72920
/feature="ced-3"
/number=100
72921..73276
/feature="ced-3"
/number=100
73277..73632
/feature="ced-3"
/number=101
73633..73988
/feature="ced-3"
/number=101
73989..74344
/feature="ced-3"
/number=102
74345..74700
/feature="ced-3"
/number=102
74701..75056
/feature="ced-3"
/number=103
75057..75412
/feature="ced-3"
/number=103
75413..75768
/feature="ced-3"
/number=104
75769..76124
/feature="ced-3"
/number=104
76125..76480
/feature="ced-3"
/number=105
76481..76836
/feature="ced-3"
/number=105
76837..77192
/feature="ced-3"
/number=106
77193..77548
/feature="ced-3"
/number=106
77549..77904
/feature="ced-3"
/number=107
77905..78260
/feature="ced-3"
/number=107
78261..78616
/feature="ced-3"
/number=108
78617..78972
/feature="ced-3"
/number=108
78973..79328
/feature="ced-3"
/number=109
79329..79684
/feature="ced-3"
/number=109
79685..80040
/feature="ced-3"
/number=110
80041..80396
/feature="ced-3"
/number=110
80397..80752
/feature="ced-3"
/number=111
80753..81108
/feature="ced-3"
/number=111
81109..81464
/feature="ced-3"
/number=112
81465..81820
/feature="ced-3"
/number=112
81821..82176
/feature="ced-3"
/number=113
82177..82532
/feature="ced-3"
/number=113
82533..82888
/feature="ced-3"
/number=114
82889..83244
/feature="ced-3"
/number=114
83245..83600
/feature="ced-3"
/number=115
83601..83956
/feature="ced-3"
/number=115
83957..84312
/feature="ced-3"
/number=116
84313..84668
/feature="ced-3"
/number=116
84669..85024
/feature="ced-3"
/number=117
85025..85380
/feature="ced-3"
/number=117
85381..85736
/feature="ced-3"
/number=118
85737..86092
/feature="ced-3"
/number=118
86093..86448
/feature="ced-3"
/number=119
86449..86804
/feature="ced-3"
/number=119
86805..87160
/feature="ced-3"
/number=120
87161..87516
/feature="ced-3"
/number=120
87517..87872
/feature="ced-3"
/number=121
87873..88228
/feature="ced-3"
/number=121
88229..88584
/feature="ced-3"
/number=122
88585..88940
/feature="ced-3"
/number=122
88941..89296
/feature="ced-3"
/number=123
89297..89652
/feature="ced-3"
/number=123
89653..90008
/feature="ced-3"
/number=124
90009..90364
/feature="ced-3"
/number=124
90365..90720
/feature="ced-3"
/number=125
90721..91076
/feature="ced-3"
/number=125
91077..91432
/feature="ced-3"
/number=126
91433..91788
/feature="ced-3"
/number=126
91789..92144
/feature="ced-3"
/number=127
92145..92500
/feature="ced-3"
/number=127
92501..92856
/feature="ced-3"
/number=128
92857..93212
/feature="ced-3"
/number=128
93213..93568
/feature="ced-3"
/number=129
93569..93924
/feature="ced-3"
/number=129
93925..94280
/feature="ced-3"
/number=130
94281..94636
/feature="ced-3"
/number=130
94637..94992
/feature="ced-3"
/number=131
94993..95348
/feature="ced-3"
/number=131
95349..95704
/feature="ced-3"
/number=132
95705..96060
/feature="ced-3"
/number=132
96061..96416
/feature="ced-3"
/number=133
96417..96772
/feature="ced-3"
/number=133
96773..97128
/feature="ced-3"
/number=134
97129..97484
/feature="ced-3"
/number=134
97485..97840
/feature="ced-3"
/number=135
97841..98196
/feature="ced-3"
/number=135
98197..98552
/feature="ced-3"
/number=136
98553..98908
/feature="ced-3"
/number=136
98909..99264
/feature="ced-3"
/number=137
99265..99620
/feature="ced-3"
/number=137
99621..100000
/feature="ced-3"
/number=138

```

BASE COUNT 2429 a 1455 c 1271 g 2498 t

Query Match 100.0%; Score 1000; DB 21; Length 7653;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5401 ATTCAAAAAAAGTCGAATTCGATTTTGTGTTTTTGGTCCCAAAAAACCAAAA 5460
 QY 5401 ATTCAAAAAAAGTCGAATTCGATTTTGTGTTTTTGGTCCCAAAAAACCAAAA 5460
 Db 5461 AAATCAATTTTCTCAAAATACCAAAAGAACCCGAAAAAATTCAGAGCTTGTCT 5520
 QY 5461 AAATCAATTTTCTCAAAATACCAAAAGAACCCGAAAAAATTCAGAGCTTGTCT 5520
 Db 5521 AATGTAACACTGATATTTAATTTCCAGGGAATGCTCTGACAATTCGAGACTTTGGCAAC 5580
 QY 5521 AATGTAACACTGATATTTAATTTCCAGGGAATGCTCTGACAATTCGAGACTTTGGCAAC 5580

Db 5581 ACGAATCACACGGAGATTCTCGGATACTCGTGATCTATCACACGGAGAGAGATGTGA 5640
 QY 5581 ACGAATCACACGGAGATTCTCGGATACTCGTGATCTATCACACGGAGAGAGATGTGA 5640
 Db 5641 TTATTGGAGTTGATGATATACCGATTAGTACACAGAGATATATGATCTTCTCAACGGGG 5700
 QY 5641 TTATTGGAGTTGATGATATACCGATTAGTACACAGAGATATATGATCTTCTCAACGGGG 5700
 Db 5701 CAATGCTCCCCGTCGGGGAATAAGCCGAAAATCGTTTTGTGAGGCTTGTGAGGCG 5760
 QY 5701 CAATGCTCCCCGTCGGGGAATAAGCCGAAAATCGTTTTGTGAGGCTTGTGAGGCG 5760
 Db 5761 GTTCGTTTTTAAATTAATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5820
 QY 5761 GTTCGTTTTTAAATTAATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5820
 Db 5821 GTGACAATGATCCCAATTCGATTCGAGGAGTTCTCGATCTTCTCGTCTG 5880
 QY 5821 GTGACAATGATTCGATTCGAGGAGTTCTCGATCTTCTCGTCTG 5880
 Db 5881 GATGGACAATCGAGACGGGCAATTTGTTCAATTTCTTGGATGTGCGGCGCAAGTTC 5940
 QY 5881 GATGGACAATCGAGACGGGCAATTTGTTCAATTTCTTGGATGTGCGGCGCAAGTTC 5940
 Db 5941 AGTTGCAATTTAATTTCTGAATGAGAAATATCTCTCAAAAATCTAAAATAGATTTT 6000
 QY 5941 AGTTGCAATTTAATTTCTGAATGAGAAATATCTCTCAAAAATCTAAAATAGATTTT 6000
 Db 6001 ATTCAGAAAGTCCGATCGAAAAATTCGATATAATTCGAAATTTGTGATAAATGAC 6060
 QY 6001 ATTCAGAAAGTCCGATCGAAAAATTCGATATAATTCGAAATTTGTGATAAATGAC 6060
 Db 6061 AAACCAATCAGCATCGATCTCGCCACTTCATCGGATTCGTTGAAAGTGGCGGA 6120
 QY 6061 AAACCAATCAGCATCGATCTCGCCACTTCATCGGATTCGTTGAAAGTGGCGGA 6120
 Db 6121 GTGAATTCGATTTGTCGAGTTTTCAGTTTGAAGGAAATTTAAAAATCGCCTTTTCCA 6180
 QY 6121 GTGAATTCGATTTGTCGAGTTTTCAGTTTGAAGGAAATTTAAAAATCGCCTTTTCCA 6180
 Db 6181 AAATTAATAATGATTTTCAATTTTTCGAAAAATTCGAAATTTTATATATCTTT 6240
 QY 6181 AAATTAATAATGATTTTCAATTTTTCGAAAAATTCGAAATTTTATATATCTTT 6240
 Db 6241 GGAGCGAAGCCCGTCTGTAAACATTTTAAATGATAATTAATAATTTTTCAGCAA 6300
 QY 6241 GGAGCGAAGCCCGTCTGTAAACATTTTAAATGATAATTAATAATTTTTCAGCAA 6300
 Db 6301 GTGTGGAGAAAGAGCGGAGCCCAAGTGCATTCGATTCGATACCAACGACGCTCAA 6360
 QY 6301 GTGTGGAGAAAGAGCGGAGCCCAAGTGCATTCGATTCGATACCAACGACGCTCAA 6360
 Db 6361 TATGTTTCGTGGAGAAACAGTCTCGTGGATCATGTTCA 6400
 QY 6361 TATGTTTCGTGGAGAAACAGTCTCGTGGATCATGTTCA 6400

RESULT 2
 LOCUS C48D1 39908 bp DNA INV 23-NOV-1998
 DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
 ACCESSION Z81049
 NID 91627677
 VERSION Z81049.1 GI:1627677
 KEYWORDS HTG.

ORGANISM
 Caenorhabditis elegans.

REFERENCE
 AUTHORS Caenorhabditis elegans.
 TITLE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 JOURNAL Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 39908)
 Direct Submission
 Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
 jes@sanger.ac.uk or rw@nemastode.wustl.edu


```

MHVAIYGKGVGQELQHVLDNRHIPVVLATKKVGFDSDEVERELALLGVTHVCVT
KRTGPGCNTIEYLEGRARKVINRDNMISATILAHMKRLGLHYTYIGGYMFAYD
KEHPGAEKKEDEFTFGSAYSVKGFDTROMNYFNQNGWNLNVRITPLSLDRE
QPNRLNLIKIYKELEDPVSLTILPDCNMAMCMEQSGTGLNVPFISLYEVV
KIYKIVDEVTNPTSIGVETERAQHLLATKGNALDTEKIQSLAPVLSAKQSLKHFS
EMTSAAK"
gene 33477..34537
/genes="C01F1.2"
join(33477..33583,33644..34347,34410..34537)
/genes="C01F1.2"
/note="coded for by C. elegans cDNA CRESW88F; similar to
S. cerevisiae protein precursors SCOI and SCOI1."
/codon_start=1
/protein_id="AAB00716.1"
/db_xref="PID:g1330396"
/db_xref="GI:1330396"
/translation="MLRTVSLACSTANLCKNTKPVWTLASAARESDKNKGDLDLETLQ
KOLKLEILKGTITETTPKPPVDKMFNFRKQAEQAFQSRSSIFNKTVLGT
FVAGGTCLAALEYIKRLDEREKHKTACKARTIGEWELMTDNGWEGSELGNW
LLMYFTNCPCDPCDETEKMKVVEIIEAKKDATPIVPVFISVDPEDRSVARKEYC
SFEKNLKGFTCTTEOVNKVAKTFRVHSQGFRTNKQEDDDYIVDHTVMYILDPGQF
HDYKGNKRAESIANVEMKVLKQAQNRKSLNLF"
33428..338975
gene
/genes="C01F1.1"
join(33428..36214,36812..36884,37527..37950,38814..38975)
/genes="C01F1.1"
/note="coded for by C. elegans cDNA yk79f12.5; coded for
by C. elegans cDNA cmi17c10; coded for by C. elegans cDNA
yk984.5; coded for by C. elegans cDNA yk59d6.5; coded for
by C. elegans cDNA CEESL34F; coded for by C. elegans cDNA
yk984.3; coded for by C. elegans cDNA yk59d6.3; coded for
by C. elegans cDNA yk79f12.3; similar to transcription
initiation factor IIf, alpha subunit"
/codon_start=1
/protein_id="AAB00717.1"
/db_xref="PID:g1330397"
/db_xref="GI:1330397"
/translation="MSGGLKPKVPGVQSEFVSRAKESDDIRYSVMFNGMDKVDTSK
WTIDSGVIMERDNRVILSTQVQEGSEYCKAAREARKKYGROSKYRLDNO
PKWMAFTEPGRQMRGIRGGANEHADIWFLKPNQSEFAKYKVDSEWHKFLPAIT
HKTLDIDQAEQFSQRYKVNQFALKAAIONQLSADSEMTQOQRLLKIKDEASD
DSDDGDEGDDGGKAKNKKKKKPAKAKRQVDEKDDVAKYESSDGEDEGREY
LIDSDDSDSDREGVDPKQVGVAAEEGARESDSESDDLTKMLKPKPYGDKK
KGNIDIERDSSGTSVDETEKLDVFMKANKDGGGGTGGKRPPTEDSDLMKDN
LGPDAKAKPAVFEGLNEETVRRYLRRKPHTTRELLHKNGKRCGNKSKSEMTQL
ASILKAIEPNQSRGKGVKLFESLVNTIA"
BASE COUNT 12825 a 7120 c 6745 g 12497 t
ORIGIN

Query Match 7.5%; Score 75; DB 21; Length 39187;
Best Local Similarity 82.6%; Pred. No. 9.51e-23;
Matches 95; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 20882 ACCGATCAGCGTCTTCGAACATCCGCCACTGGTCTGATGGTTCAAAAGTGGCGGAGC 20941
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6063 ACCATCAGATCGTCGATCTCCGCCACTTCATCGGTTGGTTGAGGAGTGGCGGAGT 6122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 20942 GAATCGCTGATGGTGGCAATCTCATTTGGAGGGAATGAAATACGCGTTT 20996
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6123 GAATTGCTGATGGTGGCAGTTTTCAGTTAGAGGGAATTAATAATCGCCTTTT 6177
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
LOCUS CELF35D11 45211 bp DNA INV 22-JUN-1995
DEFINITION Caenorhabditis elegans cosmid F35D11.
ACCESSION U29381
NID g868214
VERSION U29381.1 GI:868214
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans strain=Bristol N2.
Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea;

```

REFERENCE

1 (bases 1 to 45211)

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Pavello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkneen, R., Smaldon, N., Smith, A., Sonhammer, E., Stauden, K., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sprout, J., and Wohldman, P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

Nature 368 (6466), 32-38 (1994)

94150718

2 (bases 1 to 45211)

Fulton, B.

The sequence of C. elegans cosmid F35D11

Unpublished (1995)

3 (bases 1 to 45211)

Waterston, R.

Direct Submission

Submitted (16-JUN-1995)

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is M03A1.3' end lies in a gap followed by the cosmid T05A7. Actual end is at 45211 of CELF35D11

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

Location/Qualifiers

1..45211

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/map="II"

/strain="Bristol N2"

7036..7929

/gene="F35D11.6"

join(7036..7048,7250..7574,7625..7753,7878..7929)

/gene="F35D11.6"

/codon_start=1

/evidence=not_experimental

/protein_id="AAA68752.1"

/db_xref="PID:g868219"

/db_xref="GI:868219"

/translation="MMFIYFLPNRVHSTFGLIELKKIEKKMKILOEKKEQEDII
FYRWCLVAVRCTSNWQSGRCRYFVRLAFHVKLVKLVNERIERIYDKMIRVG
RCCELLKESERNGAIRKQFAIFSPKTPVGVSTROFLLNTTNGCAPLNRLLTLE
FFLQKNKVFASK"

complement(8583..9028)

/gene="F35D11.5"

complement(join(8583..8780,8837..8923,8978..9028))

/gene="F35D11.5"

/codon_start=1

/evidence=not_experimental

/protein_id="AAA68751.1"

/db_xref="PID:g868218"

QY	6063	ACCAATGACGATCGATCCGCCACATTCATCGGATTGGTTTGAAGTGGCGGAGT	6122	CDS	complement(join(3569..3750,3804..3945,4851..4874)) /gene="Y17G7B.1" /note="similar to Acyl CoA binding protein" /codon_start=1 /protein_id="CAA19448.1" /db_xref="PID:el323259" /db_xref="PID:g3947596" /db_xref="GI:3947596" /translation="MPISELEKKWTGLTFTFAAEEMRLKSEPTDERLUKLYAKQA LHGDIJNEDVYPVAGDEVGRKYYAAWKSQGANSEKCRADYVAIAEEMIKKGRNIV RCKWSEVWSVDY" 11683..119658 /gene="Y17G7B.2b" 11683..119658 /gene="Y17G7B.2c" 11683..119658 /gene="Y17G7B.2a" join(11683..11753,11836..11928,11980..12304,13242..13462, 14218..14331,15865..16127,17187..17299,18228..18428, 18624..18714,19438..19658) /gene="Y17G7B.2c" /note="Y17G7B.2c" /codon_start=1 /protein_id="CAA19449.1" /db_xref="PID:el323260" /db_xref="PID:g3947597" /db_xref="GI:3947597" /translation="MRSSKGGRRQAAPKTAFTTVCYCDGKRELGSVEVVCSTCLKWF HGRCLEKFEHLSNGVPMICYFTCKOCRTAEDWKAADLVQMCVTYLATLSAER LKADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN LQADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN IFVQHNDDDLFALAENKLSLGLPHEAVKLGRIERENREPHLELPIEGPKTR GAKRRHAEAPVTGKKLAADISSTAAPNGVQIDIPFSKDNRYLTLEVDNPNVDDP AWNQNSAYVIPSFYRELLNPTVNSNDRAFOLSINGSYKFSYGRSHRGHGFHEAGK SKNGQYEVNFDDQDDSHIRIGWSQASQACVYKFSYGRSHRGHGFHEAGK KYHFGFGKQDVLGLIHLPVKKLQIPANLSEKYLPSVSHGNLISFKANVFEVQ EESADIAKTLEVMGYSVIEFFHNGKSCGKAVENIYAGAYPSISIFKSATWNLGPK FRNLPRGATGIHARADEQHQEQLSDMLYLSKEVNLDPHPPRVKREDDDDVDKIKKEI KQEI" join(11683..11753,11836..11928,11980..12304,13242..13462, 14218..14331,14622..14673,18228..18428,18624..18714, 19438..19658) /gene="Y17G7B.2c" /note="cDNA EST EMBL:D74713 comes from this gene; cDNA EST YK470e5.3 comes from this gene; cDNA EST YK470e5.5 comes from this gene" /codon_start=1 /protein_id="CAA19469.1" /db_xref="PID:el356169" /db_xref="PID:g3947617" /db_xref="GI:3947617" /translation="MRSSKGGRRQAAPKTAFTTVCYCDGKRELGSVEVVCSTCLKWF HGRCLEKFEHLSNGVPMICYFTCKOCRTAEDWKAADLVQMCVTYLATLSAER LKADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN IFVQHNDDDLFALAENKLSLGLPHEAVKLGRIERENREPHLELPIEGPKTR GAKRRHAEAPVTGKKLAADISSTAAPNGVQIDIPFSKDNRYLTLEVDNPNVDDP IPFIFLASLYIEIFQGDVLGLIHLPVKKLQIPANLSEKYLPSVSHGNLISFK ANFYVEQESADIAKTLEVMGYSVIEFFHNGKSCGKAVENIYAGAYPSISIFKSAT ATWNLGPKFRNLPRGATGIHARADEQHQEQLSDMLYLSKEVNLDPHPPRVKREDDDD VKDIKKEIKQEI" join(11683..11753,11836..11928,11974..12304,13242..13462, 14218..14331,15865..16127,17187..17299,18228..18428, 18624..18714,19438..19658) /gene="Y17G7B.2b" /note="cDNA EST EMBL:D35727 comes from this gene" /codon_start=1 /protein_id="CAA19468.1" /db_xref="PID:el323261" /db_xref="PID:g3947616" /db_xref="GI:3947616" /translation="MRSSKGGRRQAAPKTAFTTVCYCDGKRELGSVEVVCSTCLKWF HGRCLEKFEHLSNGVPMICYFTCKOCRTAEDWKAADLVQMCVTYLATLSAER LKADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN IFVQHNDDDLFALAENKLSLGLPHEAVKLGRIERENREPHLELPIEGPKTR GAKRRHAEAPVTGKKLAADISSTAAPNGVQIDIPFSKDNRYLTLEVDNPNVDDP IPFIFLASLYIEIFQGDVLGLIHLPVKKLQIPANLSEKYLPSVSHGNLISFK ANFYVEQESADIAKTLEVMGYSVIEFFHNGKSCGKAVENIYAGAYPSISIFKSAT ATWNLGPKFRNLPRGATGIHARADEQHQEQLSDMLYLSKEVNLDPHPPRVKREDDDD VKDIKKEIKQEI"
QY	6123	GAATGCTGATTGGTCGCAATTCCTCATTTTGGAGGAATTTAAATAATCGCCTTT	6177	CDS	join(11683..11753,11836..11928,11980..12304,13242..13462, 14218..14331,15865..16127,17187..17299,18228..18428, 18624..18714,19438..19658) /gene="Y17G7B.2c" /note="Y17G7B.2c" /codon_start=1 /protein_id="CAA19449.1" /db_xref="PID:el323260" /db_xref="PID:g3947597" /db_xref="GI:3947597" /translation="MRSSKGGRRQAAPKTAFTTVCYCDGKRELGSVEVVCSTCLKWF HGRCLEKFEHLSNGVPMICYFTCKOCRTAEDWKAADLVQMCVTYLATLSAER LKADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN LQADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN IFVQHNDDDLFALAENKLSLGLPHEAVKLGRIERENREPHLELPIEGPKTR GAKRRHAEAPVTGKKLAADISSTAAPNGVQIDIPFSKDNRYLTLEVDNPNVDDP AWNQNSAYVIPSFYRELLNPTVNSNDRAFOLSINGSYKFSYGRSHRGHGFHEAGK SKNGQYEVNFDDQDDSHIRIGWSQASQACVYKFSYGRSHRGHGFHEAGK KYHFGFGKQDVLGLIHLPVKKLQIPANLSEKYLPSVSHGNLISFKANVFEVQ EESADIAKTLEVMGYSVIEFFHNGKSCGKAVENIYAGAYPSISIFKSATWNLGPK FRNLPRGATGIHARADEQHQEQLSDMLYLSKEVNLDPHPPRVKREDDDDVDKIKKEI KQEI" join(11683..11753,11836..11928,11980..12304,13242..13462, 14218..14331,14622..14673,18228..18428,18624..18714, 19438..19658) /gene="Y17G7B.2c" /note="cDNA EST EMBL:D74713 comes from this gene; cDNA EST YK470e5.3 comes from this gene; cDNA EST YK470e5.5 comes from this gene" /codon_start=1 /protein_id="CAA19469.1" /db_xref="PID:el356169" /db_xref="PID:g3947617" /db_xref="GI:3947617" /translation="MRSSKGGRRQAAPKTAFTTVCYCDGKRELGSVEVVCSTCLKWF HGRCLEKFEHLSNGVPMICYFTCKOCRTAEDWKAADLVQMCVTYLATLSAER LKADGKLSAEHVPEDTYLSLKDIEVPMYNNWYMLTAIKQKWHQNLAPTLLKERN IFVQHNDDDLFALAENKLSLGLPHEAVKLGRIERENREPHLELPIEGPKTR GAKRRHAEAPVTGKKLAADISSTAAPNGVQIDIPFSKDNRYLTLEVDNPNVDDP IPFIFLASLYIEIFQGDVLGLIHLPVKKLQIPANLSEKYLPSVSHGNLISFK ANFYVEQESADIAKTLEVMGYSVIEFFHNGKSCGKAVENIYAGAYPSISIFKSAT ATWNLGPKFRNLPRGATGIHARADEQHQEQLSDMLYLSKEVNLDPHPPRVKREDDDD VKDIKKEIKQEI"
Db	25417	GAATGCTGATTGGTCGCAATTCCTCATTTTGGAGGAATTTAAATAATCGCCTTT	25471		
QY	6123	GAATGCTGATTGGTCGCAATTCCTCATTTTGGAGGAATTTAAATAATCGCCTTT	6177		
RESULT	8	CEV17G7B 143092 bp DNA INV 30-NOV-1998			
LOCUS		Caenorhabditis elegans cosmid Y17G7B, complete sequence.			
DEFINITION		AL023828			
ACCESSION		93217816			
NID		HTG.			
VERSION		AL023828.1 GI:3217816			
KEYWORDS					
SOURCE		Caenorhabditis elegans.			
ORGANISM		Caenorhabditis elegans.			
REFERENCE		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;			
AUTHORS		Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
TITLE		1 (bases 1 to 143092)			
JOURNAL		Smye, R.			
REFERENCE		Direct Submission			
AUTHORS		Submitted (09-JUN-1998) Louis, MO 63110, USA. E-mail:			
TITLE		jesesanger.ac.uk or rw@nemastode.wustl.edu			
JOURNAL		2 (bases 1 to 14309			

RESULT	11	CEY45F3A	36654 bp	DNA	INV	23-NOV-1998
OCUS		Caenorhabditis elegans	cosmid	Y45F3A	complete sequence.	
DEFINITION		AL032621				
ACCESSION		g3810700				
ID		AL032621.1	GI:3810700			
VERSION						
EYWORDS		Caenorhabditis elegans.				
SOURCE		Caenorhabditis elegans				
ORGANISM		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;				
REFERENCE		Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.				
AUTHORS		1 (bases 1 to 36654)				
TITLE		Steward C.				
JOURNAL		Direct Submission				
REFERENCE		Submitted (29-OCT-1998) Louis, MO 63110, USA. E-mail: jesus@sanger.ac.uk or rwenematode.wustl.edu				
		2 (bases 1 to 36654)				

FEATURES	source
gene	
CDS	
gene	
CDS	

Downloaded from <http://ajph.org/> on November 10, 2015

```

Best Local Similarity 82.2%; Pred. No. 1.49e-19;
Matches 88; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 19269 TTTTTCAGTTCTCTCAAAATCAGAACCTCGGACCAATCAGCAATAGTCCGCCAATTTT 19328
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 6168 TTTTAAATTCCTCTAAATCGCAACTCGGACCAATCAGCAATTCACCTCGGCCACATTT 6109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19329 TCAACCAATCAGAGGAGTGGCGGAGTTCGAAGACGCTGATGGTT 19375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 6108 CAACCAATCCGATGAAGTGGCGGAGATCGAGATGCTGATGGTT 6062
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
LOCUS CEY55D9 115355 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y55D9, WORKING DRAFT SEQUENCE.
ACCESSION AL008876
NID 93378114
VERSION AL008876.1 GI:3378114
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 115355)
AUTHORS Wallis J.
JOURNAL Direct Submission
TITLES Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu
On Aug 4, 1998 this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1..115355
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y55D9"

BASE COUNT 37765 a 19764 c 20592 g 37232 t 2 others

ORIGIN
Query Match 6.9%; Score 69; DB 19; Length 115355;
Best Local Similarity 82.2%; Pred. No. 1.49e-19;
Matches 88; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 38035 AACCAATCAGCGTCTTCAAACTCGGCCACTCCGCTCATTTGGTTGAAATTCGCGGAG 38094
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6062 AACCAATCAGCATCGTCATCTCGGCCACTTCATCGGATTTGGTTTGAAGTGGCGGAG 6121
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38095 CTAATCGCTGATGGTCCAGTCTCTCATTTTGGAGGAAACTGAAAAA 38141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6122 TGAATGCTGATGGTCCAGTCTTCATTTTGAAGGAACTGAAAAA 6168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
LOCUS CEY56A3 224746 bp DNA HTG 30-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y56A3, WORKING DRAFT SEQUENCE.
ACCESSION AL022280
NID 94090202
VERSION AL022280.1 GI:4090202
KEYWORDS HTG; HTGS_PHASE1.

```

```

SOURCE
ORGANISM Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 224746)
AUTHORS Sulston J.
TITLES Direct Submission
JOURNAL Submitted (30-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu
On Dec 31, 1998 this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1..224746
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y56A3"

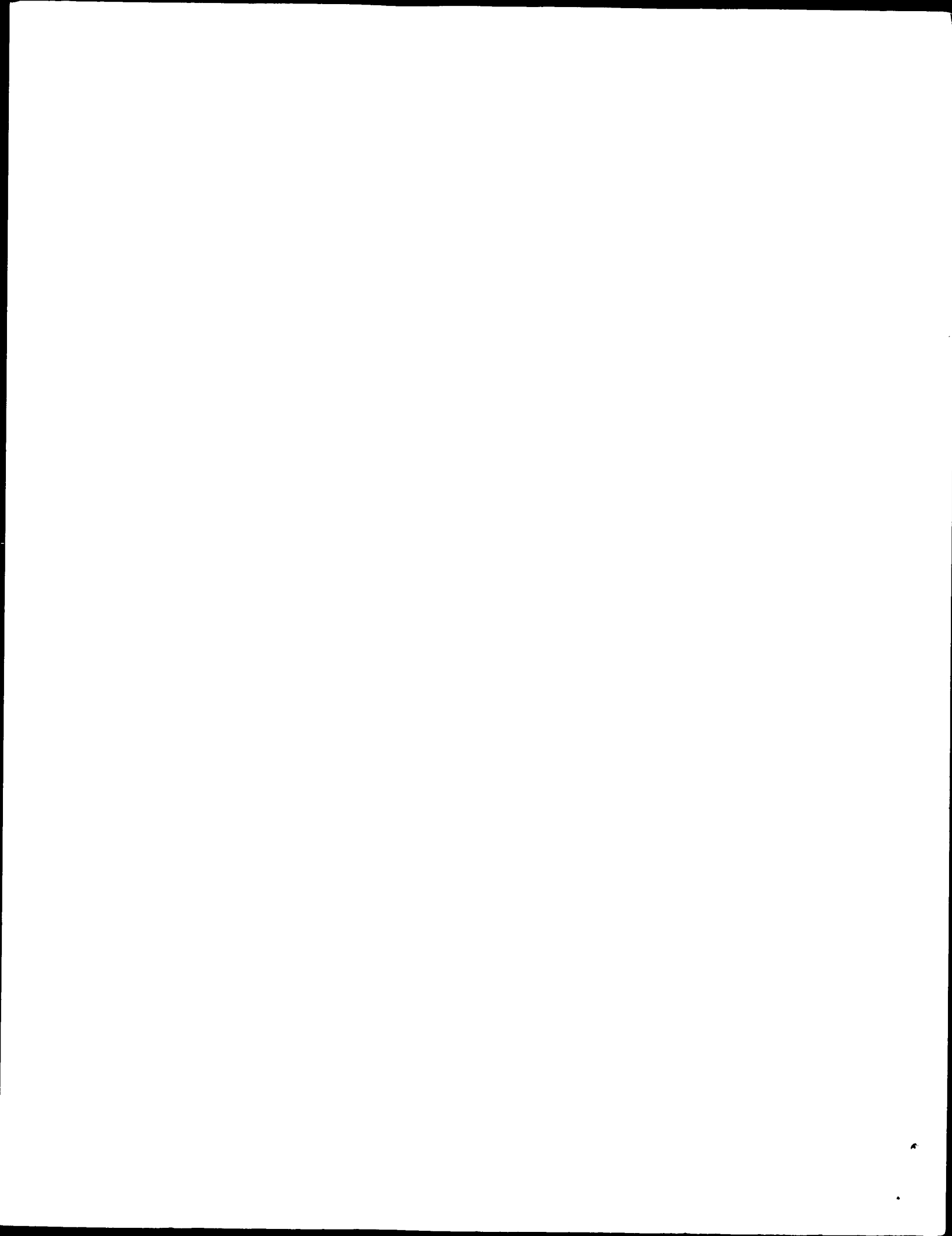
BASE COUNT 67718 a 40130 c 41166 g 69255 t 6477 others

ORIGIN
Query Match 6.9%; Score 69; DB 19; Length 224746;
Best Local Similarity 80.3%; Pred. No. 1.49e-19;
Matches 98; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

Db 129083 TAAATTCCTCCAAATCAGAACTCGGACCAATCAGCGATTCGCGCCCATTTTAAA 129142
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 6164 TAAATTCCTCTTAAACTGAAACTCGGACCAATCAGCAATTCACCTCCGCCACATTTCAA 6105
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129143 CCAATCAGATGAAGTAGGCGGAGTTTGAAGTT-CTTATTTTGAAGTGAATTTAGCCAAA 129201
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 6104 CCAATCCGATGAAGTGGCGGAGATCGAGATGCTGATTTGTTGTCATTTATCACAAA 6045
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129202 TT 129203
||
Cp 6044 TT 6043

Search completed: Sat Aug 7 21:00:35 1999
Job time : 1849 secs.

```



W P E R L H

(TW)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Aug 7 21:30:12 1999; MasPar time 235.72 Seconds
Tabular output not generated.

Title: >US-08-287-669-18
Description: (5401-6400) from US08287669.seq (7 of 10)
Perfect Score: 1000
N.A. Sequence: 5401 ATTCAAAAAAGTCGAAT.....TGTCTGTCGATCATGTTCA 6400
Comp: TAAGTTTTTTTTCAGCTTA.....ACGAGCACCTAGTACCAAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.059; Variance 6.976; scale 1.299

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1000	100.0	7653 24	T38196 Nematode Ced-3 gene.
2	1000	100.0	7653 9	Q64742 ced-3 (G6372A) gene.
3	1000	100.0	7653 9	Q64744 ced-3 (G6536A) gene.
4	1000	100.0	7653 9	Q64743 ced-3 (G6485T) gene.
5	1000	100.0	7653 9	Q64736 ced-3 (G2487A) gene.
6	1000	100.0	7653 9	Q54666 ced-3 gene.
7	1000	100.0	7653 9	Q64745 ced-3 (C7020T) gene.
8	998	99.8	7653 9	Q54401 ced-3 gene.
9	998	99.8	7653 9	Q64737 ced-3 (G5757A) gene.

10	998	99.8	7653 9	Q64740 ced-3 (C6322T) gene.
11	998	99.8	7653 9	Q64741 ced-3 (G6372A) gene.
12	998	99.8	7653 9	Q64739 ced-3 (G6297A) gene.
13	998	99.8	7653 9	Q64738 ced-3 (C5940T) gene.
14	215	21.5	2485 14	Q79970 Interleukin-1 beta co
15	43	4.3	91 46	V44650 Mammalian DNA replica
16	43	4.3	91 9	Q51746 Oligonucleotide probe
17	43	4.3	204 1	N81164 Base substituted E.co
18	43	4.3	204 1	N81164 Oligonucleotide E.co
19	37	3.7	91 9	Q51746 Generic DNA sequence
20	34	3.4	114 12	Q70468 Generic DNA sequence
21	34	3.4	114 12	Q70466 Generic DNA sequence
22	33	3.3	114 12	Q70465 Generic DNA sequence
23	33	3.3	114 12	Q70469 Generic DNA sequence
24	32	3.2	114 12	Q70468 Generic DNA sequence
25	32	3.2	114 12	Q70467 Generic DNA sequence
26	32	3.2	114 12	Q70471 Generic DNA sequence
27	32	3.2	114 12	Q70465 Generic DNA sequence
28	32	3.2	501 3	N50026 Sequence encoding new
29	31	3.1	114 12	Q70467 Generic DNA sequence
30	30	3.0	114 12	Q70466 Generic DNA sequence
31	30	3.0	114 12	Q70469 Generic DNA sequence
32	30	3.0	114 12	Q70472 Generic DNA sequence
33	30	3.0	7653 9	Q64744 ced-3 (G6536A) gene.
34	30	3.0	7653 9	Q64745 ced-3 (C7020T) gene.
35	30	3.0	7653 9	Q54401 ced-3 gene.
36	30	3.0	7653 9	Q64738 ced-3 (C5940T) gene.
37	30	3.0	7653 9	Q64742 ced-3 (G6434T) gene.
38	30	3.0	7653 9	Q64740 ced-3 (C6322T) gene.
39	30	3.0	7653 9	Q64737 ced-3 (G5757A) gene.
40	30	3.0	7653 9	Q54666 ced-3 gene.
41	30	3.0	7653 9	Q64736 ced-3 (G2487A) gene.
42	30	3.0	7653 9	Q64743 ced-3 (G6485T) gene.
43	30	3.0	7653 9	Q64741 ced-3 (G6372A) gene.
44	30	3.0	7653 9	Q64739 ced-3 (G6297A) gene.
45	30	3.0	7653 24	T38196 Nematode Ced-3 gene.

ALIGNMENTS

RESULT 1
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /*label= Repeat-1
FT repeat_unit 1490..1614
FT /*tag= b
FT /*label= Repeat-2
FT exon 2167..2366
FT /*tag= c
FT /*codon_start= 2232..2234
FT allele 2310
FT /*tag= d
FT /*label= T(n1040)
FT /*note= "causes L27F mutation"
FT intron 2367..2429
FT /*tag= e
FT /*label= Intron-1
FT exon 2430..2575
FT /*tag= f
FT allele 2487
FT /*tag= g
FT /*label= A(n718)
FT /*note= "causes G65R mutation"
FT intron 2576..2853


```

Db 5821 qtgacaatgattccagctcttgattctgtcgacgagttcgtcattcttctgctg 5880
QY 5821 GTGACAATGGATCCAGCTCTTGGAATCTGTGCGAGGAGTCCCTGCAATTCCTGCTG 5880
Db 5881 gatggacaatcgacagcgccattgttcaattttcttggatgtgtgcgcgcaagttc 5940
QY 5881 GATGGACAATCGAGAGCGGCCATTGTTCAATTTCTTGGATGTGTGGCGCGCAAGTTC 5940
Db 5941 aggttgcataattcttgaatgagaatattcctcaaaaaatcctaaatagattttt 6000
QY 5941 AGGTTGCAATTTAAATTTCTTGAATGAGATAATTCCTTCAAAAAATCTAAATAGATTTT 6000
Db 6001 attccagaagtcctcgatcgaaaaattgcatataattacgaaaatttggataaaatgac 6060
QY 6001 ATTCGAGAAGTCCCGATCGAAAAATTCGATATTAATTAACGAAATTTGTGATAAATGAC 6060
Db 6061 aaaccaatcagcatedtcgatctccgccacttcacatcgcattcggattgggttgaagtggcgga 6120
QY 6061 AAACCAATCAGCATCTCGATCTCCGCCACTTCATCGGATTGGTTTGAAGTGGCGGA 6120
Db 6121 gtgaattgctgattgtcagcttttcagtttagagggaatttaaaatcgccttttcca 6180
QY 6121 GTGAATTGCTGATTGTGCGAGTTTTCAGTTTAGAGGGAATTAATAATCGCCTTTCCA 6180
Db 6181 aaataaaattgatttttcaattttctcgaataatttcgaaaaatattccgattattttatattcttt 6240
QY 6181 AAATTAATAATTGATTTTTTCAATTTTTTCGAAAAATATTCGATTATTTTATATCTTT 6240
Db 6241 gggcgaaagcccgctcctgttaaacatttttaataataataataatttttgcagcaa 6300
QY 6241 GGAGCGAAGCCCGCTCCTGTAACATTTTAAATGATAATTAATAATTTTTCAGCAA 6300
Db 6301 gtgtgagaagaagcgcagcgaagctgacattctgattcgtacacgacacgacgtcaa 6360
QY 6301 GTGTGGAAGAAGAGCGAGCCAGCTGACATTCGTGATTCGATACGACGACGCTCAA 6360
Db 6361 tatgttctgtggagaacagctgctggtgcatcgtggtca 6400
QY 6361 TATGTTTCTGTGGAAACAGTGTGCTGTGGATCATGTGTTCA 6400

RESULT 2
ID Q64742 standard: DNA; 7653 BP.
AC Q64742;
DE 23-JUN-1994 (first entry)
KW ced-3 (C6434T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit /*tag= a
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2322..2366
FT /*tag= c
FT /*number= Exon_1
FT intron 2367..2429
FT /*tag= d
FT /*number= Intron_1
FT exon 2450..2575
FT /*tag= e
FT /*number= Exon_2
FT intron 2576..2853
FT /*tag= f
FT /*number= Intron_2

```



```

QY 6061 AAACCAATCAGCATCGATCCGCCCATTCATCGATTGGTTGAAAGTGGCGGA 6120
Db 6121 ggaattgctgattgctcagtttccagtttagagggaattaaataatcgcttttcca 6180
QY 6121 GTGAATTGCTGATTGCTCAGTTTTCAGTTTAGAGGGAATTTAAATAATCGCTTTTCCA 6180
Db 6181 aaattaaattgatttttttcaatttttcgaaattatccgattatttatatttttt 6240
QY 6181 AAATTAAAATTAATTTTTCATATTTTCGAAAAATATTCGGATATTTATATATCTTT 6240
Db 6241 ggaagcaaaagcccgctctgttaaacatttttaaatgataataataatttttgcagcaa 6300
QY 6241 GGAGCAAAAGCCCGCTCTGTAACATTTTAAATGATAATTAATAATTTTTCAGCAA 6300
Db 6301 ggtgagaaaagaagccgagcaagctgacattctgattcgatcgacacagctcaa 6360
QY 6301 GTGTGAGAAAGAACGCGAGCCAGCTGACATTCGATTCGATAGCAACGACAGCTCAA 6360
Db 6361 tatgtttcgtgagaaacagtgctgctggtgatcatgtttca 6400
QY 6361 TATGTTTCGTGAGAAACAGTCTGCTGGATCATGTTCA 6400

```

RESULT 4

```

ID O64743 standard; DNA; 7653 BP.
AC O64743;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6485T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT 2576..2853
FT /tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /tag= g
FT /number= Exon_3
FT 3108..4302
FT /tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"

```

```

FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /tag= m
FT /number= Exon_4
FT 4635..5546
FT /tag= n
FT /number= Intron_4
FT 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT 5761..5814
FT /tag= r
FT /number= Intron_5
FT 5815..5942
FT /tag= s
FT /number= Exon_6
FT 5943..6297
FT /tag= t
FT /number= Intron_6
FT 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT 6485
FT /tag= w
FT /note= "C>T, from allele n2430"
FT 6538..7012
FT /tag= x
FT /number= Intron_7
FT 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT W09325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI: 94-007542/01.
FT P-PSDB: R53286.
FT Isolated C.elegans cell death genes ced-3 and ced-4 - used to
FT developed agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT The sequences given in O64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT start codon contain repetitive elements, some of which have been
FT characterised in non-coding regions of other C. elegans genes, such
FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in

```


Db 6361 tatgttctgtgagaaacagtcgtcgatgatgttca 6400
|||||
Qy 6361 TATGTTCTGTGAGAAACAGTCTCGTGGATCATGTTCA 6400

RESULT 6
ID Q54666 standard; DNA; 7653 BP.
AC Q54666;
DT 23-JUN-1994 (first entry)
DE ced-3 gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /*tag= c
FT /number= Exon_1
FT mutation 2310
FT /*tag= d
FT /note= "C>T, from allele n1040"
FT intron 2367..2429
FT /*tag= e
FT /number= Intron_1
FT exon 2430..2575
FT /*tag= f
FT /number= Exon_2
FT mutation 2487
FT /*tag= g
FT /note= "G>A, from allele n718"
FT intron 2576..2853
FT /*tag= h
FT /number= Intron_2
FT exon 2854..3107
FT /*tag= i
FT /number= Exon_3
FT intron 3108..4302
FT /*tag= j
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /*tag= m
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /*tag= n
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /*tag= o
FT /number= Exon_4
FT intron 4635..5546
FT /*tag= p
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /*tag= q

FT repeat_unit 5221..5330
FT /*tag= r
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /*tag= s
FT /number= Exon_5
FT mutation 5757
FT /*tag= t
FT /note= "G>A, from allele n2433"
FT intron 5761..5814
FT /*tag= u
FT /number= Intron_5
FT exon 5815..5942
FT /*tag= v
FT /number= Exon_6
FT mutation 5940
FT /*tag= w
FT /note= "C>T, from allele n1165"
FT intron 5943..6297
FT /*tag= x
FT /number= Intron_6
FT repeat_region 6062..6138
FT /*tag= y
FT /rpt_type= INVERTED
FT mutation 6297
FT /*tag= z
FT /note= "G>A, from allele n717"
FT exon 6298..6537
FT /*tag= aa
FT /number= Exon_7
FT mutation 6322
FT /*tag= ab
FT /note= "C>T, from allele n1949"
FT mutation 6372
FT /*tag= ac
FT /note= "G>A, from allele n1286"
FT mutation 6434
FT /*tag= ad
FT /note= "C>T, from alleles n1129 and n1164"
FT mutation 6485
FT /*tag= ae
FT /note= "C>T, from allele n2430"
FT mutation 6535
FT /*tag= af
FT /note= "G>A, from allele n2426"
FT intron 6538..7012
FT /*tag= ag
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /*tag= ah
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /*tag= ai
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7013..7075
FT /*tag= aj
FT /number= Exon_8
FT mutation 7020
FT /*tag= ak
FT /note= "C>T, from allele n1163"
PN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.

Qy	6001	ATTCCAGAAAGTCCCGATCGAATCGAAAAATTCGGATATAAATTACGAAATTTTGATGATAAATGAC	6060
Db	6061	aaaccaatcagcatcgatctcgcgccacttcacgcgattggttgaagatgggcgga	6120
Qy	6061	AAACCAATCAGCATCGTGCATTCGCCCACTTCATCGATTGGTTTGAAGATGGCGGA	6120
Db	6121	gtgaattgctgattggtcgcagtttcagtttagaggaaattaaaaatcgccctttcga	6180
Qy	6121	GTCAATTGCTGATTGTCGCACGTTTTCAGTTTAGAGGAAATTTAAAAATCGCCTTTTCGA	6180
Db	6181	aaataaaaatgattttttcaatttttcgaaaaatattccgattatttatattcttt	6240
Qy	6181	AAATTAATAATTAATTTTTCAAATTTTTCGAAAAATATTCGGATTATTTATATATCTTT	6240
Db	6241	ggagcgaagaagccctcgtctgtaaacatttttaaatgataataataaatttttcgagcaa	6300
Qy	6241	GGAGCGAAAGCCCCGTCCTGTAAACATTTTTTAATGATAATTAATAAATTTTTTTCAGCAA	6300
Db	6301	gtgtggagaagaagccgagcccaagctgacattctgattcgatcacgcaacagctcaa	6360
Qy	6301	GTGTGGAGAAAGAGCCGAGCCCAAGCTGACATTCGTATTCGATACCGACAGCAGCTCAA	6360
Db	6361	tatgtttctggagaaacagtgcctggtgatcatggttca	6400
Qy	6361	TATGTTTCGTGGAGAAACAGTCTCGTGGATCATGGTTCA	6400

RESULT	7	
ID	Q64745	standard; DNA; 7653 BP.
AC	Q64745;	
DT	23-JUN-1994	(first entry)
DE	ced-3 (C7020t) gene.	
CE	C. elegans;	ced-4; ced-3; mutant; transcriptional regulation;
KW	embryogenesis;	cell death; hydrophilic; transmembrane; region;
KW	hydrophobic;	mutation; amino acid; substitution; RNA splicing;
KW	protein synthesis;	null phenotype; calcium-binding domain; ss.
OS	Caenorhabditis elegans.	
Key	Location/Qualifiers	
repeat_unit	1356..1472	
FT	/*tag= a	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t repeat at 1490-1614"	
FT		
repeat_unit	1490..1614	
FT	/*tag= b	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat at 1356-1472"	
FT		
exon	2232..2366	
FT	/*tag= c	
FT	/number= Exon_1	
FT	2367..2429	
FT	/*tag= d	
FT	/number= Intron_1	
FT	2450..2575	
FT	/*tag= e	
FT	/number= Exon_2	
FT	2576..2853	
FT	/*tag= f	
FT	/number= Intron_2	
FT	2854..3107	
FT	/*tag= g	
FT	/number= Exon_3	
FT	3108..4302	
FT	/*tag= h	
FT	/number= Intron_3	
FT	3126..3243	
repeat_unit	/*tag= i	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat at 3329-3396"	
FT	3329..3396	
FT	/*tag= j	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat at 3126-3243"	
FT	3487..3759	
repeat_unit		


```

UY 6121 GTGAATTGCTGATTGTCGCGAGTTTTTCAGTTTATAGAGGAATTTAAAAATCGCCTTTTCGA 6180
Db 6181 aaattaaaaattgattttttcaatttttcgaaaaattccgattattttattttattttttt 6240
Y 6181 AAATTAATAAATGATTATTTTCAATTTTTCGAAAAATATCCGATTATTTATATATCTTT 6240
Db 6241 ggaagcgaagcccgctcgtctgtaaacatttttaaatgataataataatttttgacagcaa 6300
Y 6241 GGAGCGAAGCCCGCTCTCTAAACATTTTAAATGATAATTAATAAATTTTTCAGACAA 6300
Db 6301 gtgtggagaagaagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6360
Y 6301 GTGTGGAGAAGAAGCCGAGCCAGCCAGCTGACATTCGTGATTGATACGACGACGACGCTCA 6360
Db 6361 tatgttcgtgaagaagaacagtcgtcgtggatcatggttca 6400
Y 6361 TATGTTTCGTGGAGAAACAGTCGCTGGATCATGGTTCA 6400

RESULT 12
ID Q64739 standard; DNA; 7653 BP.
AC Q64739;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6297A) gene.
KW C.elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /tag= g
FT /number= Exon_3
FT intron 3108..4302
FT /tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634

```

```

FT /tag= m
FT /number= Exon_4
FT intron 4635..5546
FT /tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT mutation 5757
FT /tag= u
FT /note= "C>T, from allele n717"
FT repeat_region 6062..6138
FT /tag= v
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= w
FT /number= Exon_7
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT W09325685-A.
FT 23-DEC-1993.
FT PD 14-JUN-1993; U05701.
FT PF 12-JUN-1992; US-897788.
FT PR 20-NOV-1992; US-979638.
FT PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT PI Horvitz HR, Shiham S, Yuan J;
FT DR WPI: 94-007542/01.
FT Isolated C.elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT start codon contain repetitive elements, some of which have been
FT characterised in non-coding regions of other C. elegans genes, such
FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT region that might be a transmembrane region. One region of Ced-3 is
FT very rich in serine. It is thought that this region is involved in
FT protein-protein interactions, similar to acid blobs in transcription
FT factors. Of the mutations which occur within the ced-3 gene, eight of
FT the mutations are missense mutations, two are nonsense mutations and

```



```

PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1: Page 42: 54pp: English.
CC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorconsensus.
CC Administration of the consensus sequence or an anti-gene (comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned
CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular or
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC elements.
SQ Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

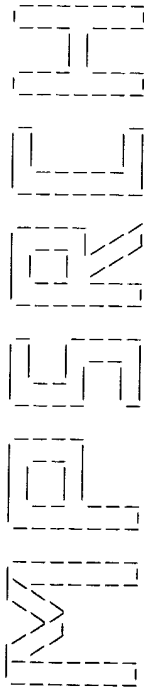
Query Match 4.3%; Score 43; DB 46; Length 91;
Best Local Similarity 18.7%; Pred. No. 1.06e-07;
Matches 17; Conservative 50; Mismatches 24; Indels 0; Gaps 0;

Db 1 awmtwaakrawrkwwkddawgwakrwkwwvhrassacmdwkaakrtwkggtwarryvkg 60
|::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Cp 5542 AAATTAATATACGATTACATTAGGAACAAAGCTGGGAATTTTTCGGGTTCTCTTTTG 5483
|::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db 61 rkmmwtwkwswdatakwwkdkakwmwrktt 91
|::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Cp 5482 GTATTTCGAGAAATGATTTTTCGGTTT 5452

Search completed: Sat Aug 7 21:34:12 1999
Job time : 240 secs.

```

RESULT					
ID	Q79970	standard; cDNA: 2485 BP.			
AC	Q79970;				
DE	Interleukin-1 beta converting enzyme homolog ced3 cDNA.				
DI	Interleukin-1 beta converting enzyme; homolog ced3;				
KW	oncogene bcl-2; programmed cell death; cancer treatment; ss.				
OS	Caenorhabditis elegans.				
FT	key	Location/Qualifiers			
FF	cds	17..1528			
ET		/*tag= a			
PB	WO9500160-A.				
PN	05-JAN-1995.				
PD	10-JUN-1994; U06630.				
PF	24-JUN-1993; US-080850.				
PP	(GENO) GEN HOSPITAL CORP.				
PI	Miura M, Yuan J;				
PT	WPI: 95-051742/07.				
DR	P-FSDB; R66770.				
DT	Promoting or preventing programmed cell death in vertebrate cells				
PT	- by inhibiting the activity of interleukin-1 beta converting				
ST	enzyme.				
PS	Example 1: Fig 2C: 116pp; English.				
CC	Q79970 encodes R6770 interleukin-1 beta converting enzyme homolog				
CC	ced3, increasing ced3's enzymatic activity can promote the				
CC	programmed cell death of cancer cells (pref. those overexpressing				
CC	the bcl-2 oncogene), this can be used as the basis of a new cancer				
CC	treatment. Alternatively by reducing ced3's enzymatic activity				
CC	programmed cell death can be inhibited, this may be useful in the				
CC	development of new cell lines which remain viable in culture for				
CC	extended or indefinite periods, independent of growth factors.				
SC	Sequence 2485 BP; 684 A; 564 C; 480 G; 757 T;				
	Query Match	21.5%; Score 215; DB 14; Length 2485;			
	Best Local Similarity	100.0%; Pred. No. 2.94e-97;			
	Matches	215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Db	883	gggaatgctcctgcacattcgagacttggccaaacogaatcatcacaggagattctgcgat	942		
Qy	5546	GGGAATGCTCTGCACAATTCGAGACTTTGCCAAACACGAATCACCGGAGATTCTCGCAT	5605		
Db	943	actcgtatcttatcacacggagaagaagtgtatttggaattgatgatataccgat	1002		
Qy	5606	ACTCGTGATTCTTATCACACGGAGAAGATGTGATTATTGGAGTTGATGATAACCGAT	5665		
Db	1003	tagtacacagatatatgatcttcacacgpgcaaatgtcccgtctgcggaataa	1062		
Qy	5666	TAGTACACACAGATATATGATCTTCTCACCGGGCAAAATGCTCCCCGTTCGCGAATAA	5725		
Db	1063	gccgaaaatcgttttgtcagaccttgcaggcg 1097			
Qy	5726	GCCGAAAATCGTTTTTGTGCAGGGCTTGTGCAGGCG 5760			
RESULT					
ID	V44650	standard; DNA: 91 BP.			
AC	V44650;				
DE	06-OCT-1998 (first entry)				
DE	Mammalian DNA replication origin consensus sequence, uniorsconsensus.				
KW	DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;				
KW	anti-gene; DNA replication inhibitor; shuttle vector construct creation				
KW	gene therapy; ss.				
OS	Mammalia.				
PN	WO9827200-A2.				
PD	25-JUN-1998.				
PF	12-DEC-1997; CA0972.				
PR	21-MAY-1997; US-047322.				
PR	16-DEC-1996; US-033374.				
PA	(UYMC-) UNIV MCGILL,				
PI	Cossens NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;				
DI	WPI: 98-36270/31.				
PT	Human or mammalian origin of replication consensus sequences - for				
PT	inhibiting DNA replication, for controlling initiation of				



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Aug 7 21:00:54 1999; Maspar time 1720.83 Seconds
Tabular output not generated. 1361.646 Million cell updates/sec

Title: >US-08-287-669-18
Description: (5401-6400) from US08287669.seq (7 of 10)
Perfect Score: 1000

N.A. Sequence: 5401 ATTCAAAAGAGTCGAAT.....TGCTGTGGATCANGTTCA 6400
Comp: TAAGTTTTTTTTCAGCTTAA.....ACGAGCACCTAGTACCAAGT

Scoring table: TABLE default
Gap 6

Mismatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_est39
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.766; Variance 3.807; scale 3.091

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	62	6-2	247	17	AA754458	97SN1784 Rice Immature
2	58	5-8	252	17	AA754459	97SN1787 Rice Immature
3	58	5-8	332	19	R03411	PK06e02.s1 Kuwabara Mi
4	57	5-7	252	17	AA754459	97SN1787 Rice Immature
5	50	5-0	360	15	C61153	C61153 Yuji Kohara unp
6	44	4-4	2275	20	AF0341173	AF0341173 Human mRNA (T
7	38	3-8	247	17	AA754458	97SN1784 Rice Immature
8	38	3-8	2275	20	AF0341173	AF0341173 Human mRNA (T
9	32	3-2	365	28	AI522052	t178b05.x1 NCI_CGAP_K1
10	31	3-1	306	28	AI570909	tr68g06.x1 NCI_CGAP_Pa

11	30	3-0	103	26	AI373276	qz48f09.x1 NCI_CGAP_K1
12	30	3-0	146	23	AI128673	ga60h01.s1 Soares_feta
13	30	3-0	230	25	AU037797	AU037797 Dictyostelium
14	30	3-0	331	18	C84704	C84704 Dictyostelium d
15	30	3-0	353	25	AU038307	AU038307 Dictyostelium d
16	30	3-0	370	22	AU049728	an37c11.x1 Gessler Wil
17	30	3-0	525	32	N46445	yy76d09.s1 Soares_mult
18	30	3-0	821	37	B21581	T25117-Sp6 TAMU Arabid
19	30	3-0	890	38	AQ026918	CIT-HSP-2322B22_TF CIT
20	29	2-9	130	25	AU037069	AU037069 Dictyostelium
21	29	2-9	221	41	AQ283940	RPC111-86M8.TV RPC111
22	29	2-9	263	21	C92260	C92260 Dictyostelium d
23	29	2-9	326	27	AI461478	496001E05.x4 496 - str
24	29	2-9	345	26	AI368191	qk65d04.x1 NCI_CGAP-Co
25	29	2-9	369	21	C90778	C90778 Dictyostelium d
26	29	2-9	371	25	AI341655	qg95e09.x1 Soares_tota
27	29	2-9	404	24	AI273546	ql55c12.x1 Soares_NHMH
28	29	2-9	414	27	AI469532	tm13f10.x1 NCI_CGAP-Co
29	29	2-9	454	33	N79808	zbl5b08.s1 Soares_feta
30	29	2-9	752	37	B21679	F8L9-Sp6 IGF Arabidops
31	29	2-9	786	26	AU003303	AU003303 Bombyx mori p
32	29	2-9	809	37	B10155	F4F14-Sp6-1 IGF Arabid
33	29	2-9	999	37	B11148	F12A9-T7 IGF Arabidops
34	28	2-8	216	35	AA039930	zk45a03.s1 Soares_preg
35	28	2-8	245	28	AI521027	to7da01.x1 NCI_CGAP_Ga
36	28	2-8	319	24	AI273112	qv64a03.x1 NCI_CGAP_Ut
37	28	2-8	327	17	AA749162	ny07f03.s1 NCI_CGAP_GC
38	28	2-8	365	22	AI040049	ox28a09.x1 Soares_tota
39	28	2-8	372	20	AA856723	od54b12.s1 NCI_CGAP_GC
40	28	2-8	377	18	AA806029	oc19h09.s1 NCI_CGAP_GC
41	28	2-8	385	18	AA761144	nz11g01.s1 NCI_CGAP_GC
42	28	2-8	395	21	AI004271	oullif04.x1 Soares_NFL
43	28	2-8	419	13	AA426444	zv47d08.s1 Soares ovar
44	28	2-8	426	37	B77427	T31F3T5B TAMU Arabidop
45	28	2-8	477	37	B40688	HS-1052-A1-E02-MR.abi

ALIGNMENTS

RESULT 1
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
CDNA clone 97SN1784, mRNA sequence.
ACCESSION AA754458
NID 92801164
VERSION AA754458.1 GI:2801164
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 247)
AUTHORS Nam, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P.,
Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y.,
Lee, M.C. and Eun, M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyungkido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..247
/organism="Oryza sativa"

/cultivar="Milyang23"
 /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional cDNA library inserted into lambda ZAPII
 vector at 5' end with EcoRI and 3' end with Xho I site."
 /db_xref="taxon:4530"
 /map="6"
 /clone="97SN1784"
 /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"

BASE COUNT 7 a 16 c 21 g 34 t 169 others
 ORIGIN

Query Match 6.2%; Score 62; DB 17; Length 247;
 Best Local Similarity 14.3%; Pred. No. 5.79e-37;
 Matches 33; Conservative 109; Mismatches 87; Indels 2; Gaps 2;
 Db 15 CBAWNNKHTMMTBWCCVRRVGTITNNGKHNRTTWNDCSDNAHCRTYVBWYARSK 74
 QY 5727 CCGAAATCGTTTGTGAGCGTTCGCGAGCGGTTCGTTTATTTAATTTAATA 5786
 Db 75 YGCTBYSNVNTGTGGTGVGKTVNVISGWNNSVYV-WETAYCDYBHYEDRAN 133
 QY 5787 TAAATATTTTAAATAATTCATTTTCAGAACGTCGTGACAAATGGATTCCAGTCTTGGAT 5846
 Db 134 HVDDT-RCINDRGYCNATSDNGTSATKRVGDKTDSDCGGCGKFKVYGSBYBRGV 192
 QY 5847 TCCTGCACGGAGTCTCTGCATTTCTTCGTCGTGGATGGACAAATCGAGACGGGCATTG 5906
 Db 193 NVVVRTSMWTKSTKMSMDMSRRSRVHYGRWBNKRGMSRNWTDTKW 243
 QY 5907 TTCAATTTCTGGATGCTGTCGGCGCCGCAAGTTCAGTTGCAATTTAATT 5957

RESULT 2
 LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
 DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
 cDNA clone 97SN1787, mRNA sequence.
 ACCESSION AA754459
 NID 92801165
 VERSION AA754459.1 GI:2801165
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euryhallophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 REFERENCE 1 (bases 1 to 252)
 AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
 Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
 Lee,M.C. and Eun,M.Y.
 Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
 Unpublished (1998)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
 Department of Cytogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyungido, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.asti.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
 University, Yongin, Korea. 449-728 bhuahm@bioserver.myongji.ac.kr
 Seq primer: M13 Reverse primer.
 Location/Qualifiers
 1..252
 /organism="Oryza sativa"
 /cultivar="Milyang23"
 /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional cDNA library inserted into lambda ZAPII
 vector at 5' end with EcoRI and 3' end with Xho I site."

FEATURES
 source

/db_xref="taxon:4530"
 /map="6"
 /clone="97SN1787"
 /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 BASE COUNT 5 a 21 c 12 g 35 t 179 others
 ORIGIN

Query Match 5.8%; Score 58; DB 17; Length 252;
 Best Local Similarity 10.9%; Pred. No. 1.30e-32;
 Matches 25; Conservative 123; Mismatches 77; Indels 5; Gaps 5;
 Db 22 SYBCHGNBVVVCVASHGNYMSVHNCBTRGTHCDCKKNVNMSTMTWGTVMNBNVSGDHHYWB 81
 Cp 5759 GCCTCGACAGCCCTGCACAAAACGATTTTCGGCTTATTCCGCAGACGGGAGCATTTGC 5700
 Db 82 VBNTKVDVGNHTRCSRWRBVT-R-MAH-Y-HDYTNCBBYNNNDHMHBBMYBBTGCMCTC 138
 Cp 5699 CGCGTTGAGAAGATCATATATCTCGTGTGTACTAATCGGTATATATCACTCAATAAT 5640
 Db 139 MWCBHYN-TKCTASGHTSTNVDKSTNTWGTBSYDKSMHGYWCSBBVKYHTKVSTT 197
 Cp 5639 CACATTCCTCTCCGTTGTATAGATCAGCAGTATCGCAGATCTCCGTTGATTCGTG 5580
 Db 198 RATRSYTCVRKY-CVMWMTKKVVKYHVBBGCBTDSKCKTMMTNKHV 246
 Cp 5579 TTTCGCAAGTCTCGAATTGTCCAGGACCATTCCTCGGAAATTAATAATCA 5530

RESULT 3
 LOCUS R03411 352 bp mRNA EST 31-MAR-1995
 DEFINITION pk06e02.s1 Kuwabara Mixed stage C. briggsae Caenorhabditis briggsae
 cDNA, mRNA sequence.
 ACCESSION R03411
 NID 9753147
 VERSION R03411.1 GI:753147
 KEYWORDS EST.
 SOURCE Caenorhabditis briggsae.
 ORGANISM Caenorhabditis briggsae.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 352)
 AUTHORS Hillier,L., Chapelli,B., Chisoe,S., Clark,N., Couch,J.,
 Dubuque,T., Hawkins,M., Holman,M., Hultman,M., Kucaba,T.,
 Kuwabara,P., Le,M., Mardis,E., Marra,M., Parsons,J., Rifkin,L.,
 Rohlfing,T., Tan,F., Trevaskis,E., Waterston,R., Wohlmann,P. and
 Wilson,R.
 Washington University Caenorhabditis briggsae EST project
 Unpublished (1995)
 COMMENT Other_ESTs: pk06e02.r1
 CONTACT: Marra MA
 Washington University Genome Sequencing Center
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1455
 Fax: 314 286 1810
 Email: mmarra@watson.wustl.edu
 PCR_F: TGTAAACGACGCGCATGAGCAAGTTCCAGCTGG
 PCR_B: CAGGAACAGCTGACCTTATGATGATTTCTCCAGGTA
 Source: Washington University Genome Sequencing Center
 PCR amplified DNA is available from Washington University Genome
 Sequencing Center. Aliquots of the library may be requested from P.
 Kuwabara (pkemrc-lmb.cam.ac.uk).
 Insert Length: 839 Std Error: 0.00
 Seq primer: Commercially available -21M13 dye primer
 High quality sequence stop: 359.
 Location/Qualifiers
 1..352
 /organism="Caenorhabditis briggsae"
 /strain="G16 Gujarat"
 /note="Vector: Lambda gt10; Site_1: EcoRI; Site_2: EcoRI;

FEATURES
 source

Stage:mixed, Sex:hermaphrodite. Library construction:
First strand oligo(GT) primed. Second strand was as per
Gubler/Hoffman. Ligated to EcoRI adaptors. Library is
non-directional. Library is non-normalized. Library
constructed by P.E. Kuwabara. Additional details on
construction of the library are described in P.E.
Kuwabara and S. Shah, NAR 22: 4414 - 4418 (1994). Adaptor
sequence: GAATTC CGTGTGTCG
/db_xref="taxon:6238"

BASE COUNT 96 a 69 c 90 g 87 t 10 others
ORIGIN

Query Match 5.8%; Score 58; DB 19; Length 352;
Best Local Similarity 70.4%; Pred. No. 1.30e-32;
Matches 126; Conservative 0; Mismatches 49; Indels 4; Gaps 4;
Db 133 ACTATCGAAGTTCCGGANGAACGATNTGCATGAGATTCTGCAATCTCGTCATTG 192
Qy 5559 ACAATTCGAGACTTTGCCAAACACGAATCACACGAGATTCTGCGTACTCGTGATCTA 5618
Db 193 TCGACGAGAGAGAGACGTTGTTATCGGAGTTGATGAGATTCTGNAACGTTTCACGA 252
Qy 5619 TCACGAGAGAGAGAGATGATTATGAGTTGATG-ATATACCGATTAGTACACACGA 5677
Db 253 NATTACGATCTCTGATGCTGCCAATGCTCTGCTGTTGGCCANTTAAGCGGAAA 311
Qy 5678 GATATATGATCTTCAACGGCGGA-ATAGTCTCC-CTCTGGCGAAT-AGCCGAAAA 5733

RESULT 4
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyungido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeunesun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea, 449-728 bnaahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
FEATURES
source
1. .252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"

/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 5.7%; Score 57; DB 17; Length 252;
Best Local Similarity 10.2%; Pred. No. 1.55e-31;
Matches 21; Conservative 108; Mismatches 75; Indels 2; Gaps 2;

Db 29 BVWCVVASHGNTMSVHNCBTRTHCDCKNVNWSMTWGTNVNBNVSGDWHYVBNVKVD 88
Qy 5609 CGTGATTCTATCACACGAGAGAATGATTGATGATGATGATGATGATGATGATGATG 5668

Db 89 VGNHTRCSRWBRVTRMAHYDYNGBYNNNDYHMHBBMYBBTGMCTMVCWCHYNTK 148
Qy 5669 TACACGAGAGATATGATCTTCAACGCGCAATGCTCCCGCTGGGGAATAAGCC 5728

Db 149 CTASGWHSTNTYDKVSSNTWGTBYSYKSMHYWCSBBVKYHTKVTSTRATRSYTCVRK 208
Qy 5729 GAAATCGTTTTTGTGAGCGCTTGTGAGCGGCTGTTTATT-TTAATT-TTAATA 5786

Db 209 YCVMWTKKVKYHVBBGCHBDS 234
Qy 5787 TAAATATTTAAATAAATCATTTTC 5812

RESULT 5
LOCUS C61153 360 bp mRNA EST 22-SEP-1997
DEFINITION C61153 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone yk225g1 5', mRNA sequence.
ACCESSION C61153
NID 92419858
VERSION C61153.1 GI:2419858
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryote; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Motonashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785439.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.

FEATURES
source
1. .360
Location/Qualifiers

/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="17q21: between D17S1321 and D17S1325"
/clone="yk225g1"
/clone_lib="Yuji Kohara unpublished cDNA"

BASE COUNT 95 a 84 c 80 g 101 t
ORIGIN

Query Match 5.0%; Score 50; DB 15; Length 360;
Best Local Similarity 82.9%; Pred. No. 3.90e-24;
Matches 63; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 43 GACCAGTCAGCAGTTTGTCTCCCGCTAGTTTGAACCAATCGGATGATGATGATGATGATG 102
Cp 6138 GACCAATACGACATTCACCTCCGCCCTTCACCAACCAATCCGATCACTGGCGGAGATC 6079

Tue Aug 10 13:03:08 1999

Contact: Tripodis, Nikos
Division of Medical and Molecular Genetics
Guys Hospital
7th floor, Guy's Tower, London SE1 9RT, UK
Email: nikos@ki.ni.

FEATURES
source

1. 2275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p21.3"
/clone="ntcon2 contig"
/clone_lib="Human mRNA (Tripodis and Ragoussis)"
438 a 619 c 470 g 599 t 149 others

BASE COUNT
ORIGIN

Query Match 3.8%; Score 38; DB 20; Length 2275;
Best Local Similarity 13.7%; Pred. No. 3.35e-12;
Matches 25; Conservative 93; Mismatches 62; Indels 3; Gaps 3;
Db 1473 AAAAAWYKRW-KRKGRKRMTGMYK-RMYRAMMAMKAMMACWYWKMRGKMKWK 1530
QY 6021 AAAAAATCGATATATACCAATTTGTGATAAATGACAAACCAATCAGCATCGTGA 6080
Db 1531 YRYKYKYSTYKSWRYWYTYTYWYWCCTSNKSGASCMRWYMGYSRYSRWGY 1590
QY 6081 TCTCGCCCACTTCATCGATGTTTGAAGAAGTGGCGGAGTGAATTCCTGATGTCGC 6140
Db 1591 WSGMSCYGMTRYRYSWTGWTKTWWYKWSMTWTMTTWTWTWTWTWTWTWTWT 1650
QY 6141 AGTTTTCAGTTTAGAGGAA-TTTAAAAATCGCCTTTTCGAAATTAATAAATGATTTT 6199
Db 1651 TCW 1653
QY 6200 TCA 6202

RESULT 9 AI522052 365 bp mRNA EST 18-MAR-1999
LOCUS ti78b05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2138097 3'
DEFINITION similar to contains element MSR1 repetitive element ;, mRNA
sequence.
ACCESSION AI522052
NID 9436187
VERSION AI522052.1 GI:4436187
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 365)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949077.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 338.
Location/Qualifiers
1..365
/organism="Homo sapiens"

FEATURES
source

/note="Organ: kidney; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo.
/db_xref="taxon:9606"
/map="MMU16C3-C4 region; MMU16C3-C4 region"
/clone="IMAGE:2138097"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
BASE COUNT 118 a 92 c 54 g 101 t

Query Match 3.2%; Score 32; DB 28; Length 365;
Best Local Similarity 71.1%; Pred. No. 9.45e-07;
Matches 54; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Db 58 TTTTNTTTTNTTTTNTTTTAAAAAAGAAAAATTTATAAGCAAAAAA 117
QY 5426 TTTTNTTTTNTTTTNTTTTGGTCCCAAAACCAAAAAATCTTCTGCAAAATACCAA 5485
Db 118 AAAAAACCCCAAAAA 133
QY 5486 AAAAAACCCCAAAAA 5501

RESULT 10 AI570909 306 bp mRNA EST 29-MAR-1999
LOCUS tr58g06.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2223514 3'
DEFINITION similar to TR:Q16038 Q16038 PRB1M PROTEIN ; contains element TARI
repetitive element ;, mRNA sequence.
ACCESSION AI570909
NID 94534283
VERSION AI570909.1 GI:4534283
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 306)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980164.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 295.
Location/Qualifiers
1..306
/organism="Homo sapiens"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 Kb. Life Technologies catalog #:
11548-013"
/db_xref="taxon:9606"
/map="11 q11-q13.4; 783C09; 5; 5p14.1-5p14.3"
/clone="IMAGE:2223514"
/clone_lib="NCI_CGAP_Pan1"

FEATURES
source

```

/tissue_type="adenocarcinoma"
/lab_host="DH10B"
BASE COUNT      70 a 63 c 60 g 112 t 1 others
ORIGIN

Query Match      3.1%; Score 31; DB 29; Length 306;
Best Local Similarity 70.1%; Pred. No. 6.88e-06;
Matches 54; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Db 58 TTTTTCGGGTTCTTTTGGTATTTTGCAGAAAAATGATTTTGGTTTGGACCA 5443
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 5502 TTTTTCGGGTTCTTTTGGTATTTTGCAGAAAAATGATTTTGGTTTGGACCA 5443
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 AAAACCAAAAAA 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 5442 AAAACCAAAAAA 5426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
LOCUS      AI373276      103 bp      mRNA      EST      16-FEB-1999
DEFINITION q248f09.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2030153 3',
            mRNA sequence.
ACCESSION  AI373276
NID        94153142
VERSION    AI373276.1 GI:4153142
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 103)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jun 15, 1998 this sequence version replaced gi:3223254.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Cloning by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 610 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 71.
Location/Qualifiers
1..103
/organism="Homo sapiens"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following hap
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855)." Subtraction by Bento Soares and M.
Fatima Bonaldo.
/db_xref="taxon:9606"
/clone="IMAGE:2030153"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
30 a 5 c 5 g 63 t

BASE COUNT      30 a 5 c 5 g 63 t
ORIGIN

Query Match      3.0%; Score 30; DB 26; Length 103;
Best Local Similarity 73.4%; Pred. No. 4.84e-05;
Matches 47; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 18 TTTTTCGGTCTTAAACCAAAAAAATTTTTCGACAAAGACAAA 77
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 5426 TTTTTCGGTCTTAAACCAAAAAAATTTTTCGACAAAGACAAA 77
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 AAAG 81
|||||

```

```

Best Local Similarity 73.4%; Pred. No. 4.84e-05;
Matches 47; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 31 TTTTTCGGTCTTAAACCAAAAAAATTTTTCGACAAAGACAAA 90
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 5426 TTTTTCGGTCTTAAACCAAAAAAATTTTTCGACAAAGACAAA 90
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91 AAAG 94
|||||
Cp 5486 AAAG 5489
|||||

RESULT 12
LOCUS      AI128673      146 bp      mRNA      EST      05-OCT-1998
DEFINITION q260H01.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1691185 3', mRNA sequence.
ACCESSION  AI128673
NID        93597187
VERSION    AI128673.1 GI:3597187
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 146)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2152106.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1138 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 93.
Location/Qualifiers
1..146
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCCATCTTTTTCCTTTTTCCTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."
/db_xref="taxon:9606"
/clone="IMAGE:1691185"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
44 a 18 c 14 g 70 t

BASE COUNT      44 a 18 c 14 g 70 t
ORIGIN

Query Match      3.0%; Score 30; DB 23; Length 146;
Best Local Similarity 73.4%; Pred. No. 4.84e-05;
Matches 47; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 18 TTTTTCGGTCTTAAACCAAAAAAATTTTTCGACAAAGACAAA 77
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 5426 TTTTTCGGTCTTAAACCAAAAAAATTTTTCGACAAAGACAAA 77
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 AAAG 81
|||||

```

LOCUS	C84704	331 bp	mRNA	EST	19-FEB-1998			
DEFINITION	C84704 Dictyostelium discoideum SS (H.Urushi-hara) Dictyostelium discoideum cDNA clone SSE61, mRNA sequence.							
ACCESSION	C84704							
NID	q2864851							
VERSION	C84704.1	GI:2864851						
KEYWORDS	EST.							
SOURCE	Dictyostelium discoideum.							
ORGANISM	Dictyostelium discoideum							
REFERENCE	Eukaryota; Dictyosteliida; Dictyostelium.							
AUTHORS	1 (bases 1 to 331)							
TITLE	Mizuno,H., Morio,T. and Tanaka,Y.							
JOURNAL	Developmental cDNA in Dictyostelium discoideum (229)							
COMMENT	Unpublished (1998)							
	On Apr 14, 1993 this sequence version replaced gi:693107.							

```

/Clone#"SSH543"
/Clone_lib#"Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage#"slug"
BASE COUNT      148 a    36 c    17 g    150 t    2 others
ORIGIN

Query Match          3.0%; Score 30; DB 25; Length 353;
Best Local Similarity 89.5%; Pred. No. 4.84e-05;
Matches 34; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db   217 TTTTTTAAAAAATTTATTTTAAATAATTAATAAAAAAAAA 254
      ||| ||||| ||||| ||||| ||||| ||||| |||||
Cp   5803 TTATTATTTAAAATATTTTATATTTAAAATTTAAAATTAATAAAAA 5766

```

Tue Aug 10 13:03:08 1999

US-08-287-669-18-07.rst

Page 8

Search completed: Sat Aug 7 21:29:52 1999
Job time : 1738 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:44:07 1999; MasPar time 23.35 Seconds
Tabular output not generated. 458.057 Million cell updates/sec

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 503
Sequence: 1 MMRQDRRSLLERNIMMFSSH.....MTRLLKKFYFWPEARNSAV 503

Scoring table: TABLE uniprotatable
Gap 60

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 2.996; Variance 0.689; scale 4.345

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	503	100.0	503 9	R47466	Ced-3.
2	503	100.0	503 19	R98754	Nematode Ced-3 cell d
3	485	96.4	503 9	R53288	Ced-3 (S486F).
4	482	95.8	503 9	R53287	Ced-3 (S483K).
5	476	94.6	503 9	R53279	Ced-3 (L27F).
6	466	92.6	503 9	R45327	Ced-3 mutant Y493.
7	466	92.6	503 9	R45326	Ced-3.
8	466	92.6	503 9	R45328	Ced-3 mutant P496.
9	466	92.6	503 9	R45326	Ced-3 mutant L488.
10	465	92.4	503 9	R45326	Ced-3 mutant N250.
11	465	92.4	503 9	R45319	Ced-3 mutant D414.
12	465	92.4	503 9	R45281	Ced-3 mutant I248.
13	465	92.4	503 9	R45294	Ced-3 mutant M291.
14	465	92.4	503 9	R45303	Ced-3 mutant G316.
15	465	92.4	503 9	R45279	Ced-3 mutant L246.
16	465	92.4	503 9	R45297	Ced-3 mutant D306.

Ced-3 mutant K349. 0.00e+00
Ced-3 mutant I438. 0.00e+00
Ced-3 mutant G261. 0.00e+00
Ced-3 mutant T287. 0.00e+00
Ced-3 mutant P344. 0.00e+00
Ced-3 mutant M234. 0.00e+00
Ced-3 mutant I247. 0.00e+00
Ced-3 mutant C358. 0.00e+00
Ced-3 mutant G277. 0.00e+00
Ced-3 mutant K283. 0.00e+00
Ced-3 mutant V280. 0.00e+00
Ced-3 mutant K351. 0.00e+00
Ced-3 mutant Q356. 0.00e+00
Ced-3 mutant D265. 0.00e+00
Ced-3 mutant S435. 0.00e+00
Ced-3 mutant F298. 0.00e+00
Ced-3 mutant N285. 0.00e+00
Ced-3 mutant R259. 0.00e+00
Ced-3 mutant S307. 0.00e+00
Ced-3 mutant P350. 0.00e+00
Ced-3 mutant S183. 0.00e+00
Ced-3 mutant H304. 0.00e+00
Ced-3 mutant F253. 0.00e+00
Ced-3 mutant A449. 0.00e+00
Ced-3 mutant L286. 0.00e+00
Ced-3 mutant G360. 0.00e+00
Ced-3 mutant I321. 0.00e+00
Ced-3 mutant H315. 0.00e+00
Ced-3 mutant R242. 0.00e+00
Ced-3 mutant D371. 0.00e+00
Ced-3 mutant L310. 0.00e+00
Ced-3 mutant S314. 0.00e+00
Ced-3 mutant V311. 0.00e+00
Ced-3 (A466V). 0.00e+00
Ced-3 mutant L346. 0.00e+00
Ced-3 mutant R439. 0.00e+00
Ced-3 mutant G323. 0.00e+00
Ced-3 mutant V452. 0.00e+00
Ced-3 mutant R359. 0.00e+00
Ced-3 mutant I334. 0.00e+00
Ced-3 mutant G434. 0.00e+00
Ced-3 mutant A357. 0.00e+00
Ced-3 mutant N339. 0.00e+00
Ced-3 mutant Y278. 0.00e+00
Nematode Ced-3 cell d 0.00e+00
Ced-3 (A449V). 0.00e+00
Ced-3 (G360S). 0.00e+00
Interleukin-1 beta co 0.00e+00
Ced-3 (G65R). 0.00e+00
Ced-3 (W428X). 0.00e+00
Ced-3 (Q412X). 0.00e+00
Ced-3 (Q403X). 0.00e+00
Ced-3. 1.01e-45
Apoptotic protease Mch 6.29e-32
Cysteine protease CMH 6.87e-02
Human interleukin-1-c 6.87e-02
Cytokine hCNTF AB loo 1.35e+00
CNTF neurotrophic pep 1.35e+00
Prosaposin receptor a 1.35e+00
Rabbit ciliary neurot 1.35e+00
Rabbit sciatic nerve 1.35e+00
Rabbit sciatic nerve 1.35e+00
Sequence of rabbit ci 1.35e+00
Modified ciliary neur 1.35e+00
Rat ciliary neurotrop 1.35e+00
Modified ciliary neur 1.35e+00
Modified ciliary neur 1.35e+00
Modified ciliary neur 1.35e+00
Amino acid sequence o 1.35e+00
Human/rat chimeric ci 1.35e+00
Human/rat chimeric ci 1.35e+00

9 R45310 503 92.4 465 17
9 R45323 503 92.4 465 18
9 R45285 503 92.4 465 19
9 R45293 503 92.4 465 20
9 R45308 503 92.4 465 21
9 R45277 503 92.4 465 22
9 R45280 503 92.4 465 23
9 R45315 503 92.4 465 24
9 R45287 503 92.4 465 25
9 R45290 503 92.4 465 26
9 R45289 503 92.4 465 27
9 R45312 503 92.4 465 28
9 R45313 503 92.4 465 29
9 R45286 503 92.4 465 30
9 R45322 503 92.4 465 31
9 R45295 503 92.4 465 32
9 R45291 503 92.4 465 33
9 R45284 503 92.4 465 34
9 R45298 503 92.4 465 35
9 R45311 503 92.4 465 36
9 R45276 503 92.4 465 37
9 R45296 503 92.4 465 38
9 R45283 503 92.4 465 39
9 R45324 503 92.4 465 40
9 R45292 503 92.4 465 41
9 R45317 503 92.4 465 42
9 R45304 503 92.4 465 43
9 R45302 503 92.4 465 44
9 R45278 503 92.4 465 45
9 R45318 503 92.4 465 46
9 R45299 503 92.4 465 47
9 R45301 503 92.4 465 48
9 R45300 503 92.4 465 49
9 R53286 503 92.4 465 50
9 R45309 503 92.4 465 51
9 R45320 503 92.4 465 52
9 R45305 503 92.4 465 53
9 R45325 503 92.4 465 54
9 R45316 503 92.4 465 55
9 R45306 503 92.4 465 56
9 R45321 503 92.4 465 57
9 R45314 503 92.4 465 58
9 R45307 503 92.4 465 59
9 R45288 503 92.4 465 60
9 R98758 503 92.4 465 61
19 R53285 503 89.9 452 62
9 R53281 503 89.9 452 63
13 R66770 503 89.9 452 64
9 R33280 503 89.9 452 65
9 R33284 427 84.9 427 66
9 R53283 411 81.7 411 67
9 R53282 402 79.9 402 68
9 R47463 505 36 7.2 69
9 R47464 480 28 5.6 70
22 W15262 303 1.8 9 71
22 W15247 303 1.8 9 72
18 R95830 341 1.8 9 73
27 W30019 20 1.6 8 74
13 R70774 20 1.6 8 75
35 W66137 20 1.6 8 76
38 W83335 20 1.6 8 77
38 W83335 199 1.6 8 78
23 R06556 199 1.6 8 79
1 R06233 199 1.6 8 80
7 R34429 199 1.6 8 81
38 W83343 200 1.6 8 82
38 W83344 200 1.6 8 83
38 W83344 200 1.6 8 84
38 W83346 200 1.6 8 85
38 W83341 200 1.6 8 86
32 W37931 200 1.6 8 87
10 R53431 200 1.6 8 88
10 R53426 200 1.6 8 89

90	1.6	200 10	R53428	Human/rat chimeric ci	1.35e+00	163	1.2	198 11	R59841	ApoB4L protease.	3.00e+02
91	1.6	200 10	R53423	Human/rat chimeric ci	1.35e+00	164	1.2	199 31	W48651	Puromycin acetyl tran	3.00e+02
92	1.6	200 10	R11617	Rat ciliary neurotrop	1.35e+00	165	1.2	200 38	W83345	Modified ciliary neur	3.00e+02
93	1.6	200 10	R53429	Human/rat chimeric ci	1.35e+00	166	1.2	200 10	R53430	Human/rat chimeric ci	3.00e+02
94	1.6	200 7	R34430	Sequence of rat cilia	1.35e+00	167	1.2	203 28	W39208	Human ICE LAP-6 amino	3.00e+02
95	1.6	821 7	R35451	Mouse eps8.	1.35e+00	168	1.2	203 28	W38184	Arabidopsis SCARPCROW	3.00e+02
96	1.4	15 16	R88457	erBB-3 peptide (aa 11	2.25e+01	169	1.2	216 15	R75909	Human olfactory recep	3.00e+02
97	1.4	15 36	W69411	Erbb-3 glycoprotein 1	2.25e+01	170	1.2	216 15	R75910	Human olfactory recep	3.00e+02
98	1.4	101 14	R77251	Guinea pig G protein	2.25e+01	171	1.2	227 31	W53835	Pseudomonas xcpw secr	3.00e+02
99	1.4	135 14	R72393	Agmenellum quadruplic	2.25e+01	172	1.2	230 23	W08430	Rapamycin-dependent a	3.00e+02
100	1.4	135 29	W55413	H. pylori ORF 29ep201	2.25e+01	173	1.2	232 37	W75092	Human secreted protei	3.00e+02
101	1.4	162 13	R67983	23kDa antigen encoded	2.25e+01	174	1.2	235 31	W27951	Amino acid sequence o	3.00e+02
102	1.4	164 38	W82399	Human G3BP protein.	2.25e+01	175	1.2	236 10	R54225	L.lactis alpha-acetol	3.00e+02
103	1.4	207 7	R36389	Urease subunit from B	2.25e+01	176	1.2	248 37	W73162	Mouse amphiregulin.	3.00e+02
104	1.4	252 23	W11617	Epstein-Barr virus p2	2.25e+01	177	1.2	249 29	W33598	Apoptosis associated	3.00e+02
105	1.4	253 22	W15263	Cysteine protease Mch	2.25e+01	178	1.2	254 12	R64190	Human 4-1BB-L polypep	3.00e+02
106	1.4	364 19	R88751	Human ty protease, re	2.25e+01	179	1.2	254 25	W26657	Human 4-1BB ligand.	3.00e+02
107	1.4	372 20	W08432	Human interleukin-1 c	2.25e+01	180	1.2	261 24	W11898	MACH isoform beta4.	3.00e+02
108	1.4	376 19	R92524	Cytoplasmic antiprote	2.25e+01	181	1.2	270 8	R42817	Class II AP endonucle	3.00e+02
109	1.4	418 15	R83098	Interleukin-1-beta co	2.25e+01	182	1.2	275 15	R84563	Human prostacyclin-sy	3.00e+02
110	1.4	567 38	W88788	Polypeptide fragment	2.25e+01	183	1.2	277 19	W06677	Pro-Yama	3.00e+02
111	1.4	667 29	W55656	H. pylori ORF 09cp610	2.25e+01	184	1.2	277 30	W41688	Amino acid sequence o	3.00e+02
112	1.4	668 30	W55738	H. pylori ORF 01ce610	2.25e+01	185	1.2	277 34	W48945	Mutant human apopain	3.00e+02
113	1.4	672 37	W73026	Helicobacter pylori 7	2.25e+01	186	1.2	277 34	W48937	Mutant human apopain	3.00e+02
114	1.4	1342 3	R13833	HER-3 epithelial grow	2.25e+01	187	1.2	277 30	W47089	Rat interleukin-1 bet	3.00e+02
115	1.4	1342 16	R88453	erBB-3 polypeptide.	2.25e+01	188	1.2	277 18	R95831	Human interleukin-I-c	3.00e+02
116	1.4	1342 36	W69406	Erbb-3 glycoprotein c	2.25e+01	189	1.2	277 24	W11893	MACH isoform beta3.	3.00e+02
117	1.4	1343 3	R12608	EGFR-R erBB-3 clone E	2.25e+01	190	1.2	277 21	W03032	Apopain CPP32beta pro	3.00e+02
118	1.4	2161 33	W63137	Human calcium channel	2.25e+01	191	1.2	277 21	W16600	Apopain CPP32a proenz	3.00e+02
119	1.4	2161 14	R71002	Human neuronal calciu	2.25e+01	192	1.2	278 37	W74863	Human secreted protei	3.00e+02
120	1.4	2161 14	R71001	Human neuronal calciu	2.25e+01	193	1.2	283 33	W49769	Amino acid sequence o	3.00e+02
121	1.4	2161 6	R33545	Sequence of the alpha	2.25e+01	194	1.2	283 21	W05302	D-aminotransferase.	3.00e+02
122	1.4	2161 33	W63149	Human calcium channel	2.25e+01	195	1.2	284 34	W75860	Human secretory prote	3.00e+02
123	1.4	3011 16	R95021	SH2 domain binding in	3.00e+02	196	1.2	292 10	R51273	SSOL gene product.	3.00e+02
124	1.2	8 34	W78520	3-phosphoinositide-de	3.00e+02	197	1.2	292 19	W06641	Infectious laryngotra	3.00e+02
125	1.2	16 36	W71727	Nck middle SH3 domain	3.00e+02	198	1.2	292 21	W06793	ILTV unique short reg	3.00e+02
126	1.2	20 27	W25350	Nck SH3 domain-bindin	3.00e+02	199	1.2	292 37	W80486	Islet cell antibody a	3.00e+02
127	1.2	20 27	W16975	Synthetic lytic pepti	3.00e+02	200	1.2	297 19	W02708	G-protein coupled hum	3.00e+02
128	1.2	23 18	R89981	Lytic peptide used in	3.00e+02	201	1.2	297 16	R48736	G-protein coupled hum	3.00e+02
129	1.2	23 18	R92424	Synthetic lytic pepti	3.00e+02	202	1.2	297 16	R48737	G-protein coupled hum	3.00e+02
130	1.2	27 18	R89982	Synthetic lytic pepti	3.00e+02	203	1.2	297 19	W02709	G-protein coupled hum	3.00e+02
131	1.2	27 18	R92425	Lytic peptide used in	3.00e+02	204	1.2	298 2	P70845	Sequence of human res	3.00e+02
132	1.2	28 39	W67932	Fragment of human sec	3.00e+02	205	1.2	298 29	W47605	HRSV glycoprotein G.	3.00e+02
133	1.2	38 7	R36838	Domain 2 from Tn554 T	3.00e+02	206	1.2	298 8	R39286	Respiratory syncytial	3.00e+02
134	1.2	46 4	R22023	Prothrombin leader fo	3.00e+02	207	1.2	298 5	R25302	HRSV glycoprotein G (3.00e+02
135	1.2	47 4	R22024	Prothrombin leader fo	3.00e+02	208	1.2	298 7	R35438	B. burgdorferi strain	3.00e+02
136	1.2	52 34	W77642	Staphylococcus aureus	3.00e+02	209	1.2	311 16	R92115	Human ApoE4Lx2.	3.00e+02
137	1.2	60 31	W57073	Measles virus haemagg	3.00e+02	210	1.2	311 11	R59843	G-protein coupled Dro	3.00e+02
138	1.2	60 31	W57087	Measles virus haemagg	3.00e+02	211	1.2	321 19	W02674	G-protein coupled Dro	3.00e+02
139	1.2	64 31	W28057	Staphylococcus aureus	3.00e+02	212	1.2	321 16	R48702	H. pylori ORF 12ap103	3.00e+02
140	1.2	73 8	R41477	Peptide derived from	3.00e+02	213	1.2	324 29	W55546	Amino acid sequence o	3.00e+02
141	1.2	78 14	R74267	SSAL.	3.00e+02	214	1.2	325 33	W59645	c-jun gene gene produ	3.00e+02
142	1.2	78 14	R74245	SSAL2 Inf A-HA-HIN1.	3.00e+02	215	1.2	334 10	R53644	HSV-2 strain SB5 Cont	3.00e+02
143	1.2	79 35	W72028	HSV-2 strain SB5 Cont	3.00e+02	216	1.2	334 36	W72171	HSV-2 strain SB5 Cont	3.00e+02
144	1.2	80 32	W49763	Human D2-2 protein as	3.00e+02	217	1.2	334 36	W72090	HSV-2 strain SB5 Cont	3.00e+02
145	1.2	87 17	R09090	Human Mat-8 polypepti	3.00e+02	218	1.2	334 35	W72005	HSV-2 strain SB5 Cont	3.00e+02
146	1.2	95 29	W55281	H. pylori ORF 12ap103	3.00e+02	219	1.2	341 16	R48727	G-protein coupled hum	3.00e+02
147	1.2	95 36	W49792	Human secreted protei	3.00e+02	220	1.2	341 19	W02699	G-protein coupled hum	3.00e+02
148	1.2	98 9	R48598	Sequence of murine IP	3.00e+02	221	1.2	344 10	R54222	L.lactis branched ami	3.00e+02
149	1.2	100 25	W17674	FRB domain of FRAP pr	3.00e+02	222	1.2	350 32	W60444	Amino acid of the spe	3.00e+02
150	1.2	110 36	W74993	Human secreted protei	3.00e+02	223	1.2	352 34	W49341	Soybean seed coat per	3.00e+02
151	1.2	111 36	W80630	S. pneumoniae spermid	3.00e+02	224	1.2	352 37	W73138	Soybean seed coat per	3.00e+02
152	1.2	111 7	R34280	Human heavy chain clo	3.00e+02	225	1.2	354 24	W25926	Xenopus melatonin rec	3.00e+02
153	1.2	121 24	R18106	Human gamma-signalin	3.00e+02	226	1.2	354 24	W25927	Xenopus melatonin rec	3.00e+02
154	1.2	124 19	W02984	G-protein coupled rec	3.00e+02	227	1.2	355 21	W14796	Mangosteen Class I ac	3.00e+02
155	1.2	124 16	W50792	G-protein coupled rec	3.00e+02	228	1.2	355 20	W09299	Mangosteen Class I ac	3.00e+02
156	1.2	130 23	W08428	pCANTAB-AP-FRAP(2015-	3.00e+02	229	1.2	356 7	R37715	Carminomycin 4-O-meth	3.00e+02
157	1.2	153 31	W54384	Actinomadura hibisca	3.00e+02	230	1.2	362 13	R74150	Brassica acyl thioest	3.00e+02
158	1.2	155 2	R12107	D.discoidium nucleosi	3.00e+02	231	1.2	362 29	W44333	Brassica campestris l	3.00e+02
159	1.2	162 28	W29461	Mad-related human JVI	3.00e+02	232	1.2	362 10	R54948	Brassica campestris l	3.00e+02
160	1.2	162 19	W04234	Mouse RAP1 partial s	3.00e+02	233	1.2	362 6	R29172	Brassica thioesterase	3.00e+02
161	1.2	195 30	W42428	Glucuronide repressor	3.00e+02	234	1.2	364 33	W57830	Rat pancreatic beta-c	3.00e+02
162	1.2	198 16	R92113	Human ApoE4L.	3.00e+02	235	1.2	370 19	W02669	G-protein coupled rat	3.00e+02

528	5	1.0	90	2	R11783	Enzyme donor polypept	2.99e+03	601	5	1.0	216	38	W89013	Polypeptide fragment	2.99e+03
529	5	1.0	92	39	W95477	Mouse derived RT3 pha	2.99e+03	602	5	1.0	216	28	W34203	Streptomyces non-memb	2.99e+03
530	5	1.0	100	38	W89015	Polypeptide fragment	2.99e+03	603	5	1.0	219	38	W88556	Secreted protein enco	2.99e+03
531	5	1.0	101	7	R34268	Mutated human VH6 gen	2.99e+03	604	5	1.0	219	22	W20650	H. pylori surface or	2.99e+03
532	5	1.0	104	39	W95223	PT toxin beta-subunit	2.99e+03	605	5	1.0	225	17	R94615	MAGE protein.	2.99e+03
533	5	1.0	105	38	W83132	Nicotiana excelsior t	2.99e+03	606	5	1.0	225	2	P70061	Sequence of human gro	2.99e+03
534	5	1.0	105	10	R53961	Hepatocyte growth fac	2.99e+03	607	5	1.0	231	13	R70000	OpDE 29 kDa subunit.	2.99e+03
535	5	1.0	107	19	W01299	VL region of HIV neut	2.99e+03	608	5	1.0	231	38	R87489	S. cerevisiae TIH3 po	2.99e+03
536	5	1.0	108	38	W85919	Amino acid sequence o	2.99e+03	609	5	1.0	232	7	R37805	P. neopolactyla typ	2.99e+03
537	5	1.0	108	17	W07440	Anti-DNA antibody 4b2	2.99e+03	610	5	1.0	234	39	W82694	Human Fc polypeptide	2.99e+03
538	5	1.0	109	22	R88753	scFv U7.6 VL.	2.99e+03	611	5	1.0	235	39	W89234	Tumour necrosis inh1b	2.99e+03
539	5	1.0	110	4	R21287	Murine VL kappa group	2.99e+03	612	5	1.0	241	5	R25245	Chicken GH receptor c	2.99e+03
540	5	1.0	125	5	F20029	Sequence of preprosom	2.99e+03	613	5	1.0	241	39	W82283	Tumour surface protei	2.99e+03
541	5	1.0	128	20	W01404	Generic ob protein.	2.99e+03	614	5	1.0	243	38	W39170	Human PKD1 protein fr	2.99e+03
542	5	1.0	130	38	W73291	R6 monoclonal antibod	2.99e+03	615	5	1.0	245	28	R60522	Anti-TNF diaphy DB32	2.99e+03
543	5	1.0	132	13	R71973	BMP-6.	2.99e+03	616	5	1.0	246	11	R60522	Human glial growth fa	2.99e+03
544	5	1.0	137	3	R38317	Sequence encoded by t	2.99e+03	617	5	1.0	248	17	R87445	Human prion protein (2.99e+03
545	5	1.0	139	39	W84214	Bone morphogenetic pr	2.99e+03	618	5	1.0	253	38	W85901	Natural killer cell p	2.99e+03
546	5	1.0	140	38	W89358	Antigen 2 from cluste	2.99e+03	619	5	1.0	254	38	W85900	Mouse prion protein (2.99e+03
547	5	1.0	140	38	W89957	Antigen 1 from cluste	2.99e+03	620	5	1.0	254	17	R94720	PRRSV VR 2385 ORF-3 p	2.99e+03
548	5	1.0	141	24	W22640	Type II topoisomerase	2.99e+03	621	5	1.0	255	38	W67571	Arabidopsis LEAFY pro	2.99e+03
549	5	1.0	143	9	R49565	Fusion protein JBI22.	2.99e+03	622	5	1.0	255	38	W85903	Sheep prion protein (2.99e+03
550	5	1.0	144	1	R80892	V region of H chain o	2.99e+03	623	5	1.0	256	38	W89870	Antigen from cluster	2.99e+03
551	5	1.0	145	38	W80288	Agroclybe aegerita fru	2.99e+03	624	5	1.0	256	38	W89992	Expressed antigen for	2.99e+03
552	5	1.0	145	3	R15323	IL-2 chimeric antibod	2.99e+03	625	5	1.0	259	31	W46998	Human telomerase reve	2.99e+03
553	5	1.0	149	38	R22648	Ehrlichia sp. extende	2.99e+03	626	5	1.0	260	38	W85902	Antigen from cluster	2.99e+03
554	5	1.0	151	38	W89809	Protein encoded by c	2.99e+03	627	5	1.0	263	38	W85902	Bovine prion protein	2.99e+03
555	5	1.0	151	38	W81777	Human HEI protein.	2.99e+03	628	5	1.0	269	14	R87475	Subtilisin N76D/I107V	2.99e+03
556	5	1.0	152	38	W89856	Antigen 2 from cluste	2.99e+03	629	5	1.0	269	39	W82802	Human soluble somatog	2.99e+03
557	5	1.0	154	38	W88690	Secreted protein enco	2.99e+03	630	5	1.0	272	13	R68346	Hepatitis C virus (HC	2.99e+03
558	5	1.0	155	38	R82620	Ehrlichia sp. JBI18.	2.99e+03	631	5	1.0	272	20	W01686	HCV protease clone C3	2.99e+03
559	5	1.0	157	9	R49564	GP4 protein of PRRSV	2.99e+03	632	5	1.0	276	38	W88934	Polypeptide fragment	2.99e+03
560	5	1.0	174	39	W84334	GP4 protein of PRRSV	2.99e+03	633	5	1.0	279	21	W02352	Proteinase K variant	2.99e+03
561	5	1.0	175	39	W84333	Fragment of human sec	2.99e+03	634	5	1.0	280	8	R42055	Insecticidal/Fungicid	2.99e+03
562	5	1.0	177	39	W68002	GP4 protein of PRRSV	2.99e+03	635	5	1.0	281	32	W60776	Amino acids 400 to 68	2.99e+03
563	5	1.0	178	39	W84328	GP4 protein of PRRSV	2.99e+03	636	5	1.0	283	2	P70392	Alpha-subunit of Sarc	2.99e+03
564	5	1.0	178	39	W86328	Kidney injury associa	2.99e+03	637	5	1.0	283	1	P90601	Fibronectin-like poly	2.99e+03
565	5	1.0	178	39	W84331	GP4 protein of PRRSV	2.99e+03	638	5	1.0	283	38	W86337	Kidney injury associa	2.99e+03
566	5	1.0	178	39	W84329	GP4 protein of PRRSV	2.99e+03	639	5	1.0	290	39	W83329	Human mindin-related	2.99e+03
567	5	1.0	178	39	W84324	GP4 protein of PRRSV	2.99e+03	640	5	1.0	290	39	W89923	Antigen from cluster	2.99e+03
568	5	1.0	178	39	W84330	GP4 protein of PRRSV	2.99e+03	641	5	1.0	291	38	W86275	Epichlorohydrin epoxi	2.99e+03
569	5	1.0	178	39	W84325	GP4 protein of PRRSV	2.99e+03	642	5	1.0	294	8	R41513	Mycobacterium fortuit	2.99e+03
570	5	1.0	178	39	W84326	GP4 protein of PRRSV	2.99e+03	643	5	1.0	298	33	W61379	Human junctional adhe	2.99e+03
571	5	1.0	178	39	W84325	GP4 protein of PRRSV	2.99e+03	644	5	1.0	298	38	W85457	Secreted protein enco	2.99e+03
572	5	1.0	178	39	W84327	GP4 protein of PRRSV	2.99e+03	645	5	1.0	299	39	W89198	Aspartate-specific cy	2.99e+03
573	5	1.0	178	7	R36537	ICE 20KD subunit.	2.99e+03	646	5	1.0	302	5	R28167	Arabidopsis lysozyme/	2.99e+03
574	5	1.0	183	39	W89689	Murine osteogenic pro	2.99e+03	647	5	1.0	303	38	W88847	Polypeptide fragment	2.99e+03
575	5	1.0	183	5	R25244	Truncated chicken GH	2.99e+03	648	5	1.0	303	25	W23380	Mycobacterium tubercu	2.99e+03
576	5	1.0	184	39	W67906	Human secreted protei	2.99e+03	649	5	1.0	310	39	W95197	Human syndecan-1.	2.99e+03
577	5	1.0	186	38	W73230	Bovine interferon-beta	2.99e+03	650	5	1.0	310	17	R70000	Human syndecan-1.	2.99e+03
578	5	1.0	186	38	W83336	Mouse ciliary neurotr	2.99e+03	651	5	1.0	314	28	W34216	Streptomyces cyclase/	2.99e+03
579	5	1.0	186	38	W83901	Bovine interferon-beta	2.99e+03	652	5	1.0	314	19	W02659	G-protein coupled dog	2.99e+03
580	5	1.0	186	38	W83900	Bovine interferon-beta	2.99e+03	653	5	1.0	316	39	W89783	Staphylococcus aureus	2.99e+03
581	5	1.0	186	38	W83905	Feline interferon-gam	2.99e+03	654	5	1.0	319	39	W87701	A human membrane fusi	2.99e+03
582	5	1.0	186	38	W73229	Bovine interferon-beta	2.99e+03	655	5	1.0	320	38	W86262	Amino acid sequence o	2.99e+03
583	5	1.0	186	38	W73234	Feline interferon-beta	2.99e+03	656	5	1.0	322	2	R24297	Glycopeptide resistan	2.99e+03
584	5	1.0	187	38	W83903	Porcine interferon-be	2.99e+03	657	5	1.0	326	39	W82311	Human A1 adenosine re	2.99e+03
585	5	1.0	187	38	W73232	Porcine interferon-be	2.99e+03	658	5	1.0	326	16	R87655	Human secreted protei	2.99e+03
586	5	1.0	189	38	W83902	Porcine interferon-al	2.99e+03	659	5	1.0	328	39	W67852	Human adenosine recep	2.99e+03
587	5	1.0	189	38	W73231	Porcine interferon-al	2.99e+03	660	5	1.0	328	38	W86336	Kidney injury associa	2.99e+03
588	5	1.0	189	5	R25692	Exons XV to XVIII of	2.99e+03	661	5	1.0	329	38	R70801	Amino acid sequence o	2.99e+03
589	5	1.0	189	6	R30611	Babesia bovis immunor	2.99e+03	662	5	1.0	329	18	R91806	Human immunoglobulin	2.99e+03
590	5	1.0	191	38	W85579	Human N-methyl-D-aspa	2.99e+03	663	5	1.0	330	5	R29643	pCTD ORF 8.	2.99e+03
591	5	1.0	191	7	R34432	Sequence of growth pr	2.99e+03	664	5	1.0	330	3	R15146	EGF-fibronectin fusio	2.99e+03
592	5	1.0	195	7	R34432	Chicken ciliary neuro	2.99e+03	665	5	1.0	331	28	W34215	Streptomyces cyclase/	2.99e+03
593	5	1.0	198	38	W83337	Pertussis toxin (del	2.99e+03	666	5	1.0	332	38	W62660	Maize OTE mature prot	2.99e+03
594	5	1.0	203	39	W89759	Staphylococcus aureus	2.99e+03	667	5	1.0	332	3	R15145	EGF-fibronectin fusio	2.99e+03
595	5	1.0	205	22	W20268	H. pylori transporter	2.99e+03	668	5	1.0	336	39	W86309	Human DNAX toll-like	2.99e+03
596	5	1.0	209	38	W82654	Ehrlichia sp. extende	2.99e+03	669	5	1.0	339	39	W86309	Kidney injury associa	2.99e+03
597	5	1.0	210	3	R12262	HIV-1 strain OII open	2.99e+03	670	5	1.0	341	3	P83200	Sequence encoded by m	2.99e+03
598	5	1.0	214	39	W82801	Human growth hormone.	2.99e+03	671	5	1.0	342	39	W87644	A protein with water	2.99e+03
599	5	1.0	215	2	R04026	Q gene product of cDN	2.99e+03	672	5	1.0	343	39	W87715	An active acyltransfe	2.99e+03
600	5	1.0	215	1	P81190	Encoded by gene y of	2.99e+03	673	5	1.0	345	39	W87715	An active acyltransfe	2.99e+03

674	5	1.0	345	19	W02694	G-protein coupled rat	2.99e+03	747	5	1.0	448	38	W67471	Borage delta-6 desatu	2.99e+03
675	5	1.0	350	2	P70645	Pseudorabies virus gp	2.99e+03	748	5	1.0	448	3	P50132	Sequence of the predo	2.99e+03
676	5	1.0	352	39	W73676	M. tuberculosis antig	2.99e+03	749	5	1.0	450	10	R53263	M. matalensis CCK B r	2.99e+03
677	5	1.0	352	39	W73786	M. tuberculosis antig	2.99e+03	750	5	1.0	454	38	W85046	Gamma subunit of a DN	2.99e+03
678	5	1.0	357	9	R44618	Human retinol binding	2.99e+03	751	5	1.0	454	4	R25155	Fibronectin-spacer-co	2.99e+03
679	5	1.0	360	22	W19703	Melanocortin-3 recept	2.99e+03	752	5	1.0	455	28	W39171	Human PKD1 protein fr	2.99e+03
680	5	1.0	361	39	W16642	Rat adenosine kinase	2.99e+03	753	5	1.0	457	39	W67743	Ovine growth hormone	2.99e+03
681	5	1.0	362	25	W16642	Truncated HSV specifi	2.99e+03	754	5	1.0	461	10	R51002	Sequence of human tum	2.99e+03
682	5	1.0	365	39	W89557	Triticum sp. cysteine	2.99e+03	755	5	1.0	463	21	W14932	Murine anti-porcine v	2.99e+03
683	5	1.0	366	38	W86259	Maize ORE enzyme.	2.99e+03	756	5	1.0	463	8	R41877	Rat glucagon-like pep	2.99e+03
684	5	1.0	367	38	W89051	Polypeptide fragment	2.99e+03	757	5	1.0	464	38	W85045	Gamma subunit of a DN	2.99e+03
685	5	1.0	368	38	W73530	Human ATG-1639 protei	2.99e+03	758	5	1.0	467	22	W05752	Presenilin-1-1 p267s	2.99e+03
686	5	1.0	368	39	W86346	Human FRAZLEDLEL protei	2.99e+03	759	5	1.0	467	4	R22759	Reshaped CD4 antibody	2.99e+03
687	5	1.0	369	39	W89245	Sphingomonas sp. stra	2.99e+03	760	5	1.0	470	39	W83037	Anti-Fas humanised an	2.99e+03
688	5	1.0	369	17	R89276	Rat tyrosine phosphat	2.99e+03	761	5	1.0	470	39	W83036	Anti-Fas humanised an	2.99e+03
689	5	1.0	369	15	R89296	Human tyrosine phosph	2.99e+03	762	5	1.0	472	39	W89559	Triticum sp. cysteine	2.99e+03
690	5	1.0	372	39	W86347	Partial human FRAZLE	2.99e+03	763	5	1.0	472	39	W86327	Kidney injury associa	2.99e+03
691	5	1.0	372	7	R39590	Trans-acting DNA bind	2.99e+03	764	5	1.0	472	38	W85456	Secreted protein enco	2.99e+03
692	5	1.0	374	39	W88201	Alcohol dehydrogenase	2.99e+03	765	5	1.0	478	21	W14500	Aspergillus oryzae al	2.99e+03
693	5	1.0	374	39	W88205	Alcohol dehydrogenase	2.99e+03	766	5	1.0	480	38	W74814	Human secreted protei	2.99e+03
694	5	1.0	374	39	W88208	Horse liver alcohol d	2.99e+03	767	5	1.0	482	39	W87766	Arabidopsis saccharop	2.99e+03
695	5	1.0	374	39	W88207	Alcohol dehydrogenase	2.99e+03	768	5	1.0	487	4	R09359	Sequence of amylase.	2.99e+03
696	5	1.0	374	39	W88206	Alcohol dehydrogenase	2.99e+03	769	5	1.0	490	39	W89764	Staphylococcus aureus	2.99e+03
697	5	1.0	374	39	W88209	Horse liver alcohol d	2.99e+03	770	5	1.0	493	39	W83125	Pril45 Arginine spe	2.99e+03
698	5	1.0	374	39	W88202	Alcohol dehydrogenase	2.99e+03	771	5	1.0	495	1	P81325	Threonine synthetase.	2.99e+03
699	5	1.0	374	39	W88200	Alcohol dehydrogenase	2.99e+03	772	5	1.0	496	5	R27801	EPSP synthase wild ty	2.99e+03
700	5	1.0	374	39	W88203	Alcohol dehydrogenase	2.99e+03	773	5	1.0	497	13	R72376	Human auxillary cytoc	2.99e+03
701	5	1.0	374	39	W88204	Alcohol dehydrogenase	2.99e+03	774	5	1.0	497	18	R93182	Human cytochrome P450	2.99e+03
702	5	1.0	374	39	W88207	Sequence encoded by 3	2.99e+03	775	5	1.0	501	33	W44868	TGF-beta superfamily	2.99e+03
703	5	1.0	374	32	W60774	Amino acids 400 to 77	2.99e+03	776	5	1.0	502	38	W81585	Human sulphamidase pr	2.99e+03
704	5	1.0	376	38	W85054	Beta subunit of DNA p	2.99e+03	777	5	1.0	502	39	W90233	CTLA4-p97 fusion prot	2.99e+03
705	5	1.0	376	25	W35387	Murine metastatic nuc	2.99e+03	778	5	1.0	503	6	R34127	Annexin XI type I fro	2.99e+03
706	5	1.0	377	26	R64143	Gibberellin-20-oxida	2.99e+03	779	5	1.0	503	3	R34127	Annexin XI type I fro	2.99e+03
707	5	1.0	377	22	W04427	Rat farnesyl transfer	2.99e+03	780	5	1.0	504	38	W86245	Mouse ActRIIB4 recept	2.99e+03
708	5	1.0	380	32	W37938	Amino acid sequence o	2.99e+03	781	5	1.0	505	28	W39903	HPV15 L1 capsid prote	2.99e+03
709	5	1.0	380	32	P93309	Promoter of beta-isop	2.99e+03	782	5	1.0	506	1	P91491	Glycoprotein A antige	2.99e+03
710	5	1.0	380	1	P90510	Sequence of an epitop	2.99e+03	783	5	1.0	509	39	W95196	Bcl-2 interaction pro	2.99e+03
711	5	1.0	387	22	W18663	Fragmented human NF-H	2.99e+03	784	5	1.0	512	38	W86243	Mouse ActRIIB2 recept	2.99e+03
712	5	1.0	387	2	P70502	Human protein C inhib	2.99e+03	785	5	1.0	514	39	W83359	Streptococcus pneumon	2.99e+03
713	5	1.0	388	38	W73514	Human TGFbetaRII:FC p	2.99e+03	786	5	1.0	516	39	W85136	A Petunia hybrida EPS	2.99e+03
714	5	1.0	388	38	W73513	Rabbit TGFbetaRII:FC p	2.99e+03	787	5	1.0	516	39	W85137	Brassica napus EPSPS	2.99e+03
715	5	1.0	389	6	R29169	Safflower long-chain	2.99e+03	788	5	1.0	516	2	P70194	Sequence of chlorophl	2.99e+03
716	5	1.0	393	21	W13949	T284R modified human	2.99e+03	789	5	1.0	522	9	R48035	Hyaluronic acid synth	2.99e+03
717	5	1.0	394	1	R05282	Amino acid sequence o	2.99e+03	790	5	1.0	527	30	W47207	Homo sapiens tubulin-	2.99e+03
718	5	1.0	394	3	P61711	[Ile358] alphas-antit	2.99e+03	791	5	1.0	528	38	W86244	Mouse ActRIIB3 recept	2.99e+03
719	5	1.0	394	12	R67363	Alpha-i-antitrypsin m	2.99e+03	792	5	1.0	529	38	W81457	Canine ICAM-R polypep	2.99e+03
720	5	1.0	397	38	W86031	Rhizobacterae phenol d	2.99e+03	793	5	1.0	529	38	W85044	Tau subunit of a DNA	2.99e+03
721	5	1.0	397	38	W86031	Rhizobacterae phenol d	2.99e+03	794	5	1.0	530	38	W90018	Expressed antigen for	2.99e+03
722	5	1.0	398	3	R15138	Human serotonin lb re	2.99e+03	795	5	1.0	532	19	R96588	Feline herpesvirus gl	2.99e+03
723	5	1.0	400	38	W88851	Polypeptide fragment	2.99e+03	796	5	1.0	536	38	W86242	Mouse ActRIIB1 recept	2.99e+03
724	5	1.0	402	38	W81966	Human VEGF-C full len	2.99e+03	797	5	1.0	540	18	W03762	Interleukin binding f	2.99e+03
725	5	1.0	405	39	W86329	Kidney injury associa	2.99e+03	798	5	1.0	543	38	W85610	Secreted protein clon	2.99e+03
726	5	1.0	409	11	R55790	Cell death hhd protei	2.99e+03	799	5	1.0	543	39	W50908	Fat cell differentiat	2.99e+03
727	5	1.0	412	30	W44190	Herpes simplex virus-	2.99e+03	800	5	1.0	545	33	W86149	Cytophaga drobachien	2.99e+03
728	5	1.0	413	39	W81967	Human P selectin liga	2.99e+03	801	5	1.0	547	38	W81440	Human intercellular a	2.99e+03
729	5	1.0	413	39	W89678	Clostridium chauvoei	2.99e+03	802	5	1.0	548	22	W20971	H. pylori cytoplasmic	2.99e+03
730	5	1.0	418	32	W56709	Amino acid sequence o	2.99e+03	803	5	1.0	549	38	W70898	Acetyl-coenzyme A tra	2.99e+03
731	5	1.0	419	38	W86203	Human vascular endoth	2.99e+03	804	5	1.0	551	24	W27150	HMG-CoA reductase deg	2.99e+03
732	5	1.0	420	4	R20200	TGF-alpha-PE40AB.	2.99e+03	805	5	1.0	555	38	W85133	A desaturase enzyme e	2.99e+03
733	5	1.0	422	38	R06450	TGF-alpha-PE40-ab mod	2.99e+03	806	5	1.0	555	38	W84154	Human desaturase enzy	2.99e+03
734	5	1.0	422	38	W73363	Chicken PAF-AH protei	2.99e+03	807	5	1.0	562	39	W95512	Amino acid sequence o	2.99e+03
735	5	1.0	422	21	W09371	Human neuroregulin GGF2	2.99e+03	808	5	1.0	567	16	R88215	Human tgf-beta type I	2.99e+03
736	5	1.0	422	21	W09371	Human GGF2.	2.99e+03	809	5	1.0	570	39	W88278	Gracilaria lemame	2.99e+03
737	5	1.0	423	17	R31305	Transcription factor-	2.99e+03	810	5	1.0	584	23	W18062	Pasteurella haemolyti	2.99e+03
738	5	1.0	424	21	W14765	Human soluble kit lig	2.99e+03	811	5	1.0	589	39	W85474	Mouse GP88 autocrine	2.99e+03
739	5	1.0	429	39	W73631	Human secreted protei	2.99e+03	812	5	1.0	592	38	W70797	Human interleukin-6R-	2.99e+03
740	5	1.0	432	38	W82246	Bacteriophage fd-tet	2.99e+03	813	5	1.0	593	7	R34669	Ap Serotype 7.65kDa T	2.99e+03
741	5	1.0	438	39	W95500	B. subtilis YabE rela	2.99e+03	814	5	1.0	593	39	W85475	Human GP88 autocrine	2.99e+03
742	5	1.0	439	38	W82645	Ehrlichia sp. extende	2.99e+03	815	5	1.0	593	5	R05874	Hsp70 antigen of Myco	2.99e+03
743	5	1.0	440	38	W73360	Murine PAF-AH protei	2.99e+03	816	5	1.0	595	38	W86003	Anti-574 single chain	2.99e+03
744	5	1.0	443	3	R27798	EPSP synthase wild ty	2.99e+03	817	5	1.0	595	38	W87487	S. cerevisiae TIH1 po	2.99e+03
745	5	1.0	446	38	W85122	A delta-6 desaturase	2.99e+03	818	5	1.0	599	39	W95462	Anti-Tac(Fv)-PE40 fus	2.99e+03
746	5	1.0	447	38	W88693	Secreted protein enco	2.99e+03	819	5	1.0	600	39	W82660	Cauliflower L-galacto	2.99e+03

820	5	1.0	605 21	W12092	Bovine papillomavirus	2.99e+03	893	5	1.0	859 38	W70796	Human gp130-Fc-His6 a	2.99e+03
821	5	1.0	607 32	W60268	Klebsiella pneumoniae	2.99e+03	894	5	1.0	860 33	W63700	Receptor type tyrosin	2.99e+03
822	5	1.0	608 26	W14645	N. meningitidis IM216	2.99e+03	895	5	1.0	876 16	R80143	B.steartothermophilus	2.99e+03
823	5	1.0	608 38	W84155	Human desaturase enzy	2.99e+03	896	5	1.0	876 16	R80143	B.steartothermophilus	2.99e+03
824	5	1.0	608 38	W85134	A desaturase enzyme e	2.99e+03	897	5	1.0	877 39	W90082	Porcine G-protein reg	2.99e+03
825	5	1.0	608 6	R33646	Rabbit pre-pro serum	2.99e+03	898	5	1.0	878 39	W95397	IBDV VP1 protein	2.99e+03
826	5	1.0	611 39	W85599	Hexosaminidase enzyme	2.99e+03	899	5	1.0	878 39	W95397	IBDV VP1 protein	2.99e+03
827	5	1.0	615 33	W95513	Amino acid sequence o	2.99e+03	900	5	1.0	878 38	W81973	Corn pullulanase prot	2.99e+03
828	5	1.0	615 33	W95513	Protein encoded by hu	2.99e+03	901	5	1.0	878 22	W20966	H. pylori flagella-as	2.99e+03
829	5	1.0	620 33	W50907	Altermomonas carragee	2.99e+03	902	5	1.0	880 39	W90088	Human G-protein regul	2.99e+03
830	5	1.0	620 33	W67882	Human secreted protei	2.99e+03	903	5	1.0	884 18	W00160	MEI2 protein kinase p	2.99e+03
831	5	1.0	625 39	W83200	Murine osteoclast dif	2.99e+03	904	5	1.0	885 20	R99740	Human phosphodiestera	2.99e+03
832	5	1.0	625 39	W77533	MEKK3 protein.	2.99e+03	905	5	1.0	889 13	R65159	Potassium ion channel	2.99e+03
833	5	1.0	630 18	R77895	Bacterial transfeerrin	2.99e+03	906	5	1.0	890 39	W73534	MEKK4 protein.	2.99e+03
834	5	1.0	631 28	W08970	Amino acid sequence o	2.99e+03	907	5	1.0	895 28	W34258	Rat ob receptor isofo	2.99e+03
835	5	1.0	636 12	R69576	Rat NDF encoded by cl	2.99e+03	908	5	1.0	908 38	W85587	Human N-methyl-D-aspa	2.99e+03
836	5	1.0	638 1	R05589	Acetohydroxy acid syn	2.99e+03	909	5	1.0	924 38	W73511	Human neuronal ICAM-4	2.99e+03
837	5	1.0	640 39	R87762	Maize lysine ketoglut	2.99e+03	910	5	1.0	929 38	W87501	N-methyl-D-aspartate	2.99e+03
838	5	1.0	655 10	R53962	Hepatocyte growth fac	2.99e+03	911	5	1.0	933 39	W89200	Human phosphatidylcho	2.99e+03
839	5	1.0	661 39	W87556	B cell surface protei	2.99e+03	912	5	1.0	936 39	W89801	Staphylococcus aureus	2.99e+03
840	5	1.0	663 24	W29150	Dual-specific murine	2.99e+03	913	5	1.0	945 39	W73624	Human secreted protei	2.99e+03
841	5	1.0	675 20	W06449	Brassica napus FCA po	2.99e+03	914	5	1.0	948 32	W48781	Thyroid peroxidase	2.99e+03
842	5	1.0	686 13	R65267	Hepatitis C virus (HC	2.99e+03	915	5	1.0	950 39	W83319	Mouse SMAD interactin	2.99e+03
843	5	1.0	686 20	W01693	Hepatitis C virus NS3	2.99e+03	916	5	1.0	951 38	W70798	Human gp130-C-gamma-1	2.99e+03
844	5	1.0	688 21	W05316	TRP-1 protein.	2.99e+03	917	5	1.0	955 8	R42241	Human p50 protein DNA	2.99e+03
845	5	1.0	693 14	R85146	Lactoferrin protein.	2.99e+03	918	5	1.0	957 22	W20841	Human p50 protein DNA	2.99e+03
846	5	1.0	698 39	W83493	4D5 Fab molecule expr	2.99e+03	919	5	1.0	965 31	W36567	H. pylori transmembra	2.99e+03
847	5	1.0	703 38	R86023	Porcine lactoferrin (2.99e+03	920	5	1.0	967 10	R51695	Toxin TccA, encoded b	2.99e+03
848	5	1.0	708 38	R86022	Bovine lactoferrin (h	2.99e+03	921	5	1.0	970 39	W89409	PEPC.	2.99e+03
849	5	1.0	709 1	P91934	Bl antigen.	2.99e+03	922	5	1.0	973 39	W83068	Maize phosphoenolpyru	2.99e+03
850	5	1.0	711 38	R86021	Human lactoferrin (hL	2.99e+03	923	5	1.0	976 39	W85581	Human N-methyl-D-aspa	2.99e+03
851	5	1.0	713 17	R85878	WD-40 domain-contg. T	2.99e+03	924	5	1.0	979 39	W83068	TutC protein of tolu	2.99e+03
852	5	1.0	718 3	R14308	N.clavipes dragline s	2.99e+03	925	5	1.0	985 8	R42995	Glycosyltransferase.	2.99e+03
853	5	1.0	722 18	R98227	Rat neuronal protein	2.99e+03	926	5	1.0	996 38	W85032	IkappaB kinase alpha	2.99e+03
854	5	1.0	728 39	R85531	Human hepatocyte grow	2.99e+03	927	5	1.0	997 38	W85032	Green fluorescent pro	2.99e+03
855	5	1.0	728 39	W85530	Human hepatocyte grow	2.99e+03	928	5	1.0	1012 39	W85589	Human N-methyl-D-aspa	2.99e+03
856	5	1.0	728 39	W85532	Human hepatocyte grow	2.99e+03	929	5	1.0	1022 39	W87761	IBDV VP2-VP3-VP4 poly	2.99e+03
857	5	1.0	728 8	R52945	Human hepatocyte grow	2.99e+03	930	5	1.0	1039 38	W85014	Maize lysine ketoglut	2.99e+03
858	5	1.0	728 8	R40862	Competative inhibitor	2.99e+03	931	5	1.0	1045 38	W85050	Stat5-green fluoresce	2.99e+03
859	5	1.0	730 22	W13669	C-proteinase encoded	2.99e+03	932	5	1.0	1045 38	W85050	Stat5-green fluoresce	2.99e+03
860	5	1.0	742 1	R05235	Amino acid sequence o	2.99e+03	933	5	1.0	1064 39	W87759	Human DNAX toll-like	2.99e+03
861	5	1.0	745 39	W82498	Human IKK-alpha prote	2.99e+03	934	5	1.0	1066 39	W88255	Arabidopsis lysine ke	2.99e+03
862	5	1.0	745 39	W86164	Human IKK-alpha poly	2.99e+03	935	5	1.0	1070 39	W88256	Morchella costata alp	2.99e+03
863	5	1.0	745 38	W81562	IKB kinase (IKK)-alph	2.99e+03	936	5	1.0	1086 38	W81839	Morchella vulgaris alp	2.99e+03
864	5	1.0	746 38	W81565	IKB kinase (IKK)-alph	2.99e+03	937	5	1.0	1092 39	W83257	Human LFA-1 alpha cha	2.99e+03
865	5	1.0	746 38	W81565	A desaturase enzyme e	2.99e+03	938	5	1.0	1092 39	W83257	Gracilariaopsis lemame	2.99e+03
866	5	1.0	746 38	W84156	Human desaturase enzy	2.99e+03	939	5	1.0	1099 20	W05177	Lats large tumour sup	2.99e+03
867	5	1.0	752 38	W82466	Ehrlichia sp. extende	2.99e+03	940	5	1.0	1118 38	W82395	Human UBP protein #1.	2.99e+03
868	5	1.0	753 39	W95314	Amino acid sequence o	2.99e+03	941	5	1.0	1125 30	W52288	Rattus norvegicus cdo	2.99e+03
869	5	1.0	764 18	R91809	Human IgG CHI-Hinge-C	2.99e+03	942	5	1.0	1130 20	W05178	H-lats large tumour s	2.99e+03
870	5	1.0	769 32	W60795	Deletion mutant delta	2.99e+03	943	5	1.0	1132 31	W46957	Human telomerase reve	2.99e+03
871	5	1.0	770 23	W19485	APP770 mutant A-beta-	2.99e+03	944	5	1.0	1134 29	W46859	Bacillus thuringiens	2.99e+03
872	5	1.0	774 26	W35390	Flavobacterium odorat	2.99e+03	945	5	1.0	1151 38	W73344	Rat alphas protein se	2.99e+03
873	5	1.0	774 17	R93984	Cephalosporin C acyla	2.99e+03	946	5	1.0	1151 38	W73346	Mouse alphas protein se	2.99e+03
874	5	1.0	774 15	R86374	Cephalosporin C acyla	2.99e+03	947	5	1.0	1161 38	W73347	Mouse alphas protein se	2.99e+03
875	5	1.0	774 15	R86373	Cephalosporin C acyla	2.99e+03	948	5	1.0	1161 38	W73347	Green fluorescent pro	2.99e+03
876	5	1.0	777 33	W54079	Hexosaminidase enzyme	2.99e+03	949	5	1.0	1171 38	W85037	NFAT1-green fluoresce	2.99e+03
877	5	1.0	777 33	W54079	Hexosaminidase enzyme	2.99e+03	950	5	1.0	1181 38	W85038	Guanine thymine bindi	2.99e+03
878	5	1.0	784 39	W87760	Soybean lysine ketogl	2.99e+03	951	5	1.0	1184 20	W10676	Human Serum Albumin a	2.99e+03
879	5	1.0	789 29	W46861	Bacillus thuringiens	2.99e+03	952	5	1.0	1219 38	W87509	Human N-methyl-D-aspa	2.99e+03
880	5	1.0	798 38	W85025	Smad4-green fluoresce	2.99e+03	953	5	1.0	1219 39	W87507	Human N-methyl-D-aspa	2.99e+03
881	5	1.0	806 16	R86352	Human DNAX toll-like	2.99e+03	954	5	1.0	1229 39	W87533	CryET5 protein sequen	2.99e+03
882	5	1.0	809 39	W86330	Tomato pathogen resis	2.99e+03	955	5	1.0	1231 38	W87506	Human N-methyl-D-aspa	2.99e+03
883	5	1.0	806 38	W53103	Anabaena sucrose synt	2.99e+03	956	5	1.0	1236 38	W85574	Human N-methyl-D-aspa	2.99e+03
884	5	1.0	806 38	W53103	Smad4-green fluoresce	2.99e+03	957	5	1.0	1239 38	W85508	Human N-methyl-D-aspa	2.99e+03
885	5	1.0	810 28	W19521	B. cereus VIP3A(a) sy	2.99e+03	958	5	1.0	1243 38	W87505	Human N-methyl-D-aspa	2.99e+03
886	5	1.0	810 28	W19521	Mouse interleukin-4 r	2.99e+03	959	5	1.0	1247 39	W87535	Human N-methyl-D-aspa	2.99e+03
887	5	1.0	810 39	W73471	Human signal mediator	2.99e+03	960	5	1.0	1274 39	W89253	Human N-methyl-D-aspa	2.99e+03
888	5	1.0	834 22	W13506	Human DNAX toll-like	2.99e+03	961	5	1.0	1323 10	R52548	Human ALP.	2.99e+03
889	5	1.0	837 39	W86361	Human DNAX toll-like	2.99e+03	962	5	1.0	1336 38	W87511	N-methyl-D-aspartic a	2.99e+03
890	5	1.0	841 28	W46397	Amino acid sequence o	2.99e+03	963	5	1.0	1336 38	W87511	Human N-methyl-D-aspa	2.99e+03
891	5	1.0	847 39	W85603	Hexosaminidase enzyme	2.99e+03	964	5	1.0	1336 38	W87511	Tissue-binding hybrid	2.99e+03
892	5	1.0	849 38	W82397	Human UBP protein #3.	2.99e+03	965	5	1.0	1336 33	W57837	Aldehyde oxidase prot	2.99e+03

966 5 1.0 1363 2 R11061 Bovine Coronavirus E2 2.99e+03
 967 5 1.0 1451 38 W83310 LRP5 protein from iso 2.99e+03
 968 5 1.0 1454 9 R42474 FECV/FIPV chimeric sp 2.99e+03
 969 5 1.0 1479 3 R13231 CFTR delta 1507. 2.99e+03
 970 5 1.0 1480 3 R13232 CFTR G85E. 2.99e+03
 971 5 1.0 1484 38 W87510 Human N-methyl-D-aspa 2.99e+03
 972 5 1.0 1493 5 R27640 Human calcium channel 2.99e+03
 973 5 1.0 1561 18 W20298 S. mutans antigen I/I 2.99e+03
 974 5 1.0 1591 38 W83308 Toxin TccB, encoded b 2.99e+03
 975 5 1.0 1614 38 W83312 Mature LRP5 protein. 2.99e+03
 976 5 1.0 1615 38 W83311 Mouse Lrp5 protein. 2.99e+03
 977 5 1.0 1639 38 W83311 LRP5 isoform 3 protei 2.99e+03
 978 5 1.0 1706 24 W24786 PrkA antigenic protei 2.99e+03
 979 5 1.0 1763 39 W83431 Rat Munc13-1. 2.99e+03
 980 5 1.0 1852 22 W10011 Protein encoded by mu 2.99e+03
 981 5 1.0 1852 19 R99448 BRCA1 allele #77 tran 2.99e+03
 982 5 1.0 1863 18 R81539 BRCA1 mutant from PM1 2.99e+03
 983 5 1.0 1863 18 R81514 BRCA1 mutant from sam 2.99e+03
 984 5 1.0 1863 18 R81511 BRCA1 mutant from sam 2.99e+03
 985 5 1.0 1938 39 W73553 Full length Pullulana 2.99e+03
 986 5 1.0 2100 39 W89579 Calcium permeable vol 2.99e+03
 987 5 1.0 2104 39 W89578 Calcium permeable vol 2.99e+03
 988 5 1.0 2105 39 W89577 Calcium permeable vol 2.99e+03
 989 5 1.0 2150 2 R05127 Complete human rhino 2.99e+03
 990 5 1.0 2209 5 P20037 Sequence encoded by a 2.99e+03
 991 5 1.0 2324 16 R92778 Human fibronectin. 2.99e+03
 992 5 1.0 2435 5 R25135 HCV polypeptide 1. 2.99e+03
 993 5 1.0 2639 39 W73476 Grapevine leafroll vi 2.99e+03
 994 5 1.0 2766 22 W19651 ATM mutant 8283delTC. 2.99e+03
 995 5 1.0 2799 39 W81867 Human tumour suppress 2.99e+03
 996 5 1.0 2938 11 R59923 GAP protein Iral. 2.99e+03
 997 5 1.0 3033 6 R33214 NANBH virus strain HC 2.99e+03
 998 5 1.0 3085 22 W19701 ATM mutant G9170C. 2.99e+03
 999 5 1.0 3088 2 R06996 Protein characteristi 2.99e+03
 1000 5 1.0 3088 2 R06996

ALIGNMENTS

RESULT 1
 ID R47466 standard; Protein; 503 AA.
 AC R47466;
 DT 23-JUN-1994 (first entry)
 DE Ced-3.
 KW C. elegans; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain.
 CS Caenorhabditis elegans.
 PN W09325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PK 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI; 94-007542/01.
 DR N-PSDB; Q54666.
 PT Isolated C. elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PT Clam 6; Fig 4; 127pp; English.
 CS This sequence is encoded by the C. elegans ced-3 gene. A 2.8 kb mRNA
 CC was identified as the ced-3 transcript and was most abundant in
 CC embryos, but was also detected in larvae and young adults, suggesting
 CC that ced-3 is expressed not only in cells undergoing cell death. The
 CC four largest introns as well as sequences 5' of the start codon
 CC contain repetitive elements, some of which have been characterised
 CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
 CC and myoB. The Ced-3 protein is 503 amino acids in length. Ced-3 is
 CC highly hydrophilic with no significant hydrophobic region that might
 CC be a transmembrane region. One region of Ced-3 is very rich in serine.
 CC It is thought that this region is involved in protein-protein inter-
 CC actions, similar to acid blobs in transcription factors. Of the

CC mutations which occur within the ced-3 gene, eight of the mutations
 CC are missense mutations, two are nonsense mutations and two are putative
 CC splicing mutations (see also R53279-88). These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 503 AA;

Query Match 100.0%; Score 503; DB 9; Length 503;

Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrqrrrrllnnimfshlkvdeilevliakqvlinsdngdmnscgtvrekrrivka 60
 QY 1 MMQRRRLLENNIMFSSHLKVEILEVLIKAQVLSNDGDMNSCGTVREKREIVKA 60
 Db 61 vrrgdvafdafalrstgheglaevleplarsvdsnavefecpmpashrralspa 120
 QY 61 VRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSDVSNVAFECFPMSPASHRRSRALSPA 120
 Db 121 gytstprvhrdsvsvsftsyqdiysrarsrrsralshrhnyssppvnaifspqss 180
 QY 121 GYTSTPRVHRDSVSVSFTSYQDIYSRARSRSRALHSSDRHNSPPVNAFPSPSS 180
 Db 181 anssftgcsslgyssrrnrfsksgtqyifheedmfvdaptisrvfdektmyrnfs 240
 QY 181 ANSFTGCSLGYSSRRNRFSKSGTQYIFHEEDMNFVDAPTISRVEKTMRYNFS 240
 Db 241 prgmclinnhefemqtrngtkadknltnlfrcmtyvtickdnltrgmlltirfak 300
 QY 241 PRGMCLINNEHEFEMQTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRFAK 300
 Db 301 heshgdailvshgenviigvddipistheilydlalnaaprlankpkivfvqacrg 360
 QY 301 HESHGDSAILVLSHGENVIIGVDDIPISTHEIYDILNNAAPRLANKPKIVFVQACRG 360
 Db 361 errdngfpvldsvdgpafirrgwmdrdgplfnflgcvrpgvqvwrkpsqadiliry 420
 QY 361 ERRDNGFPVLDSDGVPALFRGWDNRDGPLFNLGCVRFQVQVWRKPKSQADILIRYA 420
 Db 421 ttagyvwnrsargswfiqavcevfsthakmdvelltvevknkvacgftsggsnlkq 480
 QY 421 TTAGYVWNRSARGSWFIQAVCEVFSTHAKMDVELLTVENKKNVACGFTSGGSNLKQ 480
 Db 481 mpemtsrllkkfyfwearnsav 503
 QY 481 MPEMTSRLLKKFYFWEARNSAV 503

RESULT 2

ID R98754 standard; Protein; 503 AA.
 AC R98754;
 DT 17-DEC-1996 (first entry)
 DE Nematode Ced-3 cell death protein.
 KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
 KW cell death; apoptosis; neural degeneration; inflammation;
 KW antiinflammatory.
 OS Caenorhabditis elegans.
 PH Key Location/Qualifiers
 FT region 78..205
 FT /label= Serine-rich_region
 FT cleavage_site 131
 FT /note= "autocleavage site"
 FT cleavage_site 371
 FT /note= "autocleavage site"
 FT active_site 338
 FT /note= "active site Cysteine residue"
 PN W09625946-A1.
 PD 29-AUG-1996.
 PF 23-FEB-1996; U02473.

```

PR 24-FEB-1995; US-394189.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR WPI: 96-425082/42.
DR N-PSDB; T38196.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
PT Claim 18; Fig 3; 139pp; English.
CC The Caenorhabditis elegans Ced-3 protein (R98754) is encoded
CC by the ced-3 cell death gene (T38196). It is structurally similar
CC to human interleukin-1 beta converting enzyme (ICE) (R98755); the
CC 2 proteins have 28% amino acid identity (higher in the C-terminal
CC regions critical for activity), suggesting that Ced-3 may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death protein. Ced-3 proteins can therefore
CC be used in prodn. of novel drugs for enhancing or inhibiting the
CC activity of ICE for the treatment of inflammatory diseases and/or
CC diseases caused by cell death. Novel inhibitors of ced-3 activity
CC include portions of Ced-3. Ced-3 with a C360A substn. (see also
CC R98758) can prevent programmed cell death in C. elegans.
CC Sequence 503 AA;
SQ
Query Match 100.0%; Score 503; DB 19; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 mmqrdrslernimfshlkvdlellvliakqvlnsdngminscgtvrekrrvika 60
QY 1 MMRQDRSLERNIMFSSHLKVDEILVLIKQVLNSDNGMINSCGTVEKRRRIVKA 60
Db 61 vqrgdvafdafydlrstgheglavleplarsvdsnavfecpmpashrralspa 120
QY 61 VQRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECMPASHRRALS 120
Db 121 gytsptrvhrdsvssvftsqdiysrarsrslhssdrhnyssppvnapfsgps 180
QY 121 GYTSPTVRHRDSVSSVFTSQDIYSRARSRSRSLHSSDRHNYSSPPVNAFPSP 180
Db 181 ansftgcslgyssrnrfsksagptqyifheedmfvdpaptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSRNRFSKSGPTQYIFHEEDMNFVDAPTISRVEDEKTYRNF 240
Db 241 prgmcllinnehfeompnrngtkadkdnltlncrmgytvickdnltrgmlltirdfak 300
QY 241 PRGMCLLINNEHFEOMPNRNGTKADKDNLTNLCRCMGYTVICKDNLTRGMLLT 300
Db 301 heshgdsailvshgeenvilgvdipistheiydlldnaaprlanpkivfvqacrg 360
QY 301 HESHGDSAILVSHGEEENVILGVDDIPISTHEIYDLLNAANAPRLANPKIVFVQ 360
Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpqgvvrkpsqadiliry 420
QY 361 ERRDNGFPVLDSVDGVPAPFLRRGWDNRDGPLFNFLGCVRPQGVQVVRKPSQA 420
Db 421 ttaqyvswnrsargswfioavcevfsthakdmvveltevnkkvacgftsggsnllkq 480
QY 421 TTAQYVSWNRNARGSWFIOAVCEVFSTHAKDMVVELTEVNKKVACGFTSGGSN 480
Db 481 mpemtsrllkkkfyfpearnsav 503
QY 481 MPEMTSRLKKKFYFWEARNSAV 503
RESULT
ID R53288 standard; Protein; 503 AA.
AC R53288;
DE 23-JUN-1994 (first entry)
DT Ced-3 (S486F).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.

```

```

OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 486
FT /label= S486F
FT /note= "From allele nll163"
PN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR WPI: 94-007542/01.
DR N-PSDB; Q64745.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PT Claim 13; Fig 4; 127pp; English.
PS The sequences given in R5379-88 are encoded by mutant versions of the
PS C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
PS transcript and was most abundant in embryos, but was also detected in
PS larvae and young adults, suggesting that ced-3 is expressed not only
PS in cells undergoing cell death. The four largest introns as well as
PS sequences 5' of the start codon contain repetitive elements, some of
PS which have been characterised in non-coding regions of other C. elegans
PS genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
PS acids in length. Ced-3 is highly hydrophilic with no significant
PS hydrophobic region that might be a transmembrane region. One region
PS of Ced-3 is very rich in serine. It is thought that this region is
PS involved in protein-protein interactions, similar to acid blobs in
PS transcription factors. Of the mutations which occur within the ced-3
PS gene, eight of the mutations are missense mutations, two are nonsense
PS mutations and two are putative splicing mutations. These mutations
PS establish the null phenotype of the ced-3 gene, confirming that ced-3,
PS like ced-4, function is not essential for viability. The ced-3 and
PS ced-4 gene products may be used to develop agents for treating
PS conditions characterised by cell deaths, such as myocardial infarction,
PS stroke, degenerative disease, traumatic brain injury, hypoxia,
PS pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;
Query Match 96.4%; Score 485; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 mmqrdrslernimfshlkvdlellvliakqvlnsdngminscgtvrekrrvika 60
QY 1 MMRQDRSLERNIMFSSHLKVDEILVLIKQVLNSDNGMINSCGTVEKRRRIVKA 60
Db 61 vqrgdvafdafydlrstgheglavleplarsvdsnavfecpmpashrralspa 120
QY 61 VQRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECMPASHRRALS 120
Db 121 gytsptrvhrdsvssvftsqdiysrarsrslhssdrhnyssppvnapfsgps 180
QY 121 GYTSPTVRHRDSVSSVFTSQDIYSRARSRSRSLHSSDRHNYSSPPVNAFPSP 180
Db 181 ansftgcslgyssrnrfsksagptqyifheedmfvdpaptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSRNRFSKSGPTQYIFHEEDMNFVDAPTISRVEDEKTYRNF 240
Db 241 prgmcllinnehfeompnrngtkadkdnltlncrmgytvickdnltrgmlltirdfak 300
QY 241 PRGMCLLINNEHFEOMPNRNGTKADKDNLTNLCRCMGYTVICKDNLTRGMLLT 300
Db 301 heshgdsailvshgeenvilgvdipistheiydlldnaaprlanpkivfvqacrg 360
QY 301 HESHGDSAILVSHGEEENVILGVDDIPISTHEIYDLLNAANAPRLANPKIVFVQ 360
Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpqgvvrkpsqadiliry 420
QY 361 ERRDNGFPVLDSVDGVPAPFLRRGWDNRDGPLFNFLGCVRPQGVQVVRKPSQA 420
Db 421 ttaqyvswnrsargswfioavcevfsthakdmvveltevnkkvacgftsggsnllkq 480
QY 421 TTAQYVSWNRNARGSWFIOAVCEVFSTHAKDMVVELTEVNKKVACGFTSGGSN 480

```

```

|||||
QY 421 TTAQVSVNRNSRGSWFIQAVCEVFSTHAKDMVDVVELLTVNKKVACGFTSQSGSNILKQ 480
Db 481 mpemt 485
QY 481 MPEMT 485

RESULT 4
ID R53287 standard; Protein; 503 AA.
AC R53287;
DT 23-JUN-1994 (first entry)
DE Ced-3 (E483K).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 483
FT /label= E483K
FT /note= "From allele n2426"
DN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
RN N-PSDB; Q64744.
RT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 13; Fig 4; 127pp; English.
CC The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
CC Sequence 503 AA;
SQ

Query Match 95.88; Score 482; DB 9; Length 503;
Best Local Similarity 100.08; Pred. No. 0.00e+00;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrgdrslrllernimmfshlkdleilvliakqvinsdngdmnscgtvrekrrivka 60
QY 1 MMQRDRSLRLERNIMFSSHLKVDLEILVLIKQVLSNDGDMNSCGTVREKRREIVKA 60

Db 61 vqrgdvafdafdalrstghevlaevleplarsvdsnavfecpmshrralspa 120
QY 61 VQRGDVAFDAFDFALRSTGHEGLAEVLEPLARSDVSNAVEFECPMSPASHRRALSPPA 120

Db 121 gytsptrvhrdsvssvffsyvdiyrarsrrsrllshsdrrhyssppvnafpsqps 180
QY 121 GYTSPTRVHRDSVSSVFFSYVDIYRARSRRSRLLSHSDRRHYSSPPVNAFPSPSS 180

```

```

Db 181 ansstgcseslgyssrnrfskasgptqyifheedmfvdpdaptisrvfdetkmyrnfs 240
QY 181 ANSSTGCSLSGLYSSRNRFSKASGPTQYIFHEEDMNFVDAPTISRVEDEKTYRNFS 240

Db 241 pigmclliinnehefecmptrngtkadknltnlfcmgvtyvickdnltrgmlltirdfak 300
QY 241 PRGMCLLIINNEHEFEQMPTRNGTKADKDLTNLFCMGVTVICKDNLTRGMLLTIRDFAK 300

Db 301 heshgdsailvilshgeenviigvdipistheiydlinaanaplankpkivfqacrg 360
QY 301 HESGDSAILVILSHGEENVIIIGVDIPISTHEIYDLINAANAPLANKPKIVFQACRG 360

Db 361 errdngfpvldsvdvpafllrgwdnrdgplfnflgcvrpqvgvwrkpkpsqadiliry 420
QY 361 ERDNGFPVLDSVDGVPFLLRRGWNDRDGPLFNFLGCVRPQVGQVWRKPKPSQADILIRYA 420

Db 421 ttaqvsvnrnsargswfiqavcevfsthakdmvdvveltevnkkvacgftsqsgsnilkq 480
QY 421 TTAQVSVNRNSRGSWFIQAVCEVFSTHAKDMVDVVELLTVNKKVACGFTSQSGSNILKQ 480

Db 481 mp 482
QY 481 MP 482

RESULT 5
ID R53279 standard; Protein; 503 AA.
AC R53279;
DT 23-JUN-1994 (first entry)
DE Ced-3 (L27F).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 27
FT /label= L27F
FT /note= "From allele n1040"
DN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
RN N-PSDB; Q64735.
RT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 13; Fig 4; 127pp; English.
CC The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
CC Sequence 503 AA;
SQ

```


RESULT	12		
ID	089110	PRELIMINARY:	PRT: 480 AA.
AC	089110;		
DT	01-NOV-1998	(TREMBREL, 08, CREATED)	
DT	01-NOV-1998	(TREMBREL, 08, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998	(TREMBREL, 08, LAST ANNOTATION UPDATE)	
DE	CASPASE-8.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA.		
OC	SCIDUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129 SVJ.		
RX	MEDLINE; 98316661.		
RA	SAKAMAKI K., TSUKUMO S., YONEHARA S.;		
RT	"Molecular cloning and characterization of mouse caspase-8.";		
RL	EUR. J. BIOCHEM. 253:399-405(1998).		
DR	EMBL; AF067841; G3211998; -		
DR	EMBL; AF067835; G3211998; JOINED.		
DR	EMBL; AF067836; G3211998; JOINED.		
DR	EMBL; AF067837; G3211998; JOINED.		
DR	EMBL; AF067838; G3211998; JOINED.		
DR	EMBL; AF067839; G3211998; JOINED.		
DR	EMBL; AF067840; G3211998; JOINED.		
DR	EMBL; AF067834; G3193167; -		

[illegible]

QY 305 -GDSAILVILSHGEENVYIGV--DDIP-I-STHEIYDLNANAPRLANKPRIVEVOACR 359
 Db 165 GDNIGSV-QIRDSID--PE-MPTS-TS-DG--WR---CYELGSKT-REAHLESPATLY 212
 C 360 CERNDNFPLVDSDGVPALFRGMNRODPLFNFGCVPOVOOYWRKKPSQADLLRY 419
 Db 213 SSTPDVSWSPRESFTIORLVEOFNRHA--FNS-DL-QEMFRKYORSFENP-RQ-L- 265
 QY 420 ATTAQYVSWNSRARGSMFIQAVCEVFSTAKMDVVELLEVKKYKACGFGQSQGSNLIK 479
 Db 266 PTOERT-MKKTYLP 281
 QY 480 QMPEMSTRLLKKEFPW 496

RESULT 8
 ID 093417 PRELIMINARY: PRT: 283 AA.
 AC 093417;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CASPASE-3.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA JOHNSON A.L., BRIDGHAM J.T.;
 RT "Characterization of the chicken caspase-3 cDNA."
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF083029; G3450875; -
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS01122; CASPASE_CYS; 1.
 SQ SEQUENCE 283 AA; 31675 MW; 745D181C CRC32;

Query Match 8.2%; Score 295; DB 13; Length 283;
 Best Local Similarity 33.1%; Pred. No. 5, 87e-37;
 Matches 95; Conservative 77; Mismatches 74; Indels 41; Gaps 21;

Db 27 MNLPAKSVSDGILPDDG-YRMDYPE-IGCVIYNNKNNHRTGTSRSGTADAAVRE 84
 QY 217 MNEVDAPITSR-VPEDEKTYR-NFSSPRGMCLINNEH--EQ-MPTRNGTAKDKNLTN 271
 Db 85 VEMKLGKVKLNNDLSSRDIFKLKNVS-EEDHSKRSSFVCLSHDEGLFTGDP-PL 142
 QY 272 LFRCMGYTICDKNLTGRMMLTIRDFAKHESHG-SALT-VILSHGEENVYIGVDLPI 329
 Db 143 ELKVLTLRGRDKCSLAKPKLFTIQACRGT--E-----IDS--GIDA-----DS--G 185
 QY 330 STHEIYDLNANAPRLANKPRIVEVOACRGERDNGFVLDSVDGVPALFRGMNDNRD 389
 Db 186 PD-ETV-C-----OKT---PVEADFLYASTAPGYYSRNAAESWFIQSCRLKKEHA 234
 QY 390 PLFNFGCVPOVOOYWRKKPSQADLLRYATTAQYVSWNSRARGSMFIQAVCEVFSTHA 449
 Db 235 RLKELMQLTRNRRVAYEESCTRODFNAKROICIVSMLEKEEYF 281
 QY 450 KMDVVELLEVKKYKVA-CGF-QTSGSNILKQEMPMSTRLLKKEYP 494

RESULT 9
 ID 077623 PRELIMINARY: PRT: 182 AA.
 AC 077623;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CASPASE-3 (FRAGMENT).
 OS OVIS ARIES (SHEEP).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLIA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; CARPINEAE; OVIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA RUEDA B.R., HENDRI I.R., TILLY J.L., HAMERNIK D.L.;

RT "Accumulation of Caspase-3 mRNA and Induction of Caspase Activity in
 RT the Ovine Corpus Luteum Following Prostaglandin-F2a Treatment in
 RT vivo."
 RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF068837; G3293438; -
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS01122; CASPASE_CYS; 1.
 FT NON TER
 SQ SEQUENCE 182 AA; 24048 MW; 0FD658EC CRC32;

Query Match 7.5%; Score 269; DB 6; Length 182;
 Best Local Similarity 35.1%; Pred. No. 1, 73e-31;
 Matches 46; Conservative 33; Mismatches 45; Indels 7; Gaps 6;
 Db 6 YKMDYPE-MGLCIITNNKNNHENTGMACRSGTDVDAANLRETTSLKEYRINDLTRE 64
 QY 235 YR-NFSSPRGMCLINNEH--EQ--MPTNGTAKDKNLTNFRGMGYTICKDKNLTGR 290
 Db 65 MLEIMSVSKEDHSKRSSFICVLLSHGEGLIFGTNG-PVDLKLASFPGDCRSITGK 123
 QY 291 MLTIRDFAKHESHGDSALT-VILSHGEENVYIGVDIPISTHEIYDLNANAPRLANK 349
 Db 124 PKLFIQACRG 134
 QY 350 PRIVEVOACRG 360

RESULT 10
 ID 075601 PRELIMINARY: PRT: 377 AA.
 AC 075601;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE EVOLUTIONARILY RELATED INTERLEUKIN-1BETA CONVERTING ENZYME.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98288312.
 RA HUMKE E.W., NI J., DIXIT V.M.;
 RT "ERICE, a novel FLICE-activatable caspase."
 RL J. BIOL. CHEM. 273:15702-15707(1998).
 DR EMBL: AF078533; G3386523; -
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS01122; CASPASE_CYS; 1.
 SQ SEQUENCE 377 AA; 43012 MW; CB1FDF3A3 CRC32;

Query Match 7.4%; Score 268; DB 4; Length 377;
 Best Local Similarity 37.8%; Pred. No. 2, 80e-31;
 Matches 48; Conservative 30; Mismatches 42; Indels 7; Gaps 5;

Db 139 LIICNTEFDMPPRNGAALDILGMKQLBGLGYVEEKLTAQDMESVLMKFAAREER 198
 QY 246 LIINNEHEQMPTRNGKADKDKNLTNLFRCMGYVICDKNLTGRGLTIRDFAKHESH- 304
 Db 199 SSDSTFLVEMSHGILDGICGTMSEEDPVLPTDITFRFPNNRNCISLQDKRVIIVQAC 258
 QY 305 -GDSAILVILSHGEENVYIG-V--DDIP-ISTHE-IYDLNANAPRLANKPRIVEVOAC 358

Db 259 RGANRGE 265
 QY 359 RGERDN 365
 RESULT 11
 ID 089094 PRELIMINARY: PRT: 257 AA.
 AC 089094;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CASPASE-14.
 GN CASP-14.

[illegible]

```

RESULT      5      PRELIMINARY:      PRT:      277 AA.
ID      035397
AC      035397;
DT      01-JAN-1998 (TREMBLREL. 05, CREATED)
DT      01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT      01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE      CASPASE 6.
GN      MCH2.
OS      RATIUS NORVEGICUS (RAT).
OC      EUCARYOTA; METAZOA; CHORATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA.
OC      SCURROGNATHI; MURIDAE; MURINAE; RATIUS.
[1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-SPRAGUE-DAWLEY, TISSUE-KIDNEY CORTEX;
RA      KAUSHAL G.P., SINGH A.B., SHAH S.V.;
RT      "Rat kidney Mch2.";
RL      SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR      EMBL: AF025670; G3287307; -.
DR      PROSITE: PS01121; CASPASE_HIS: 1.
DR      PROSITE: PS01122; CASPASE_CYS: 1.
SQ      SEQUENCE 277 AA; 3156 MM; 2FCOC446 CRC32;

```

Query Match	9.78;	Score 350;	DB 11;	Length 277;
Best Local Similarity	34.98;	Pred. No. 8.07e-49;		
Matches 97; Conservative	59;	Mismatches 97;	Indels 25;	Gaps 17

```

Db 10 SREVLDAEDQYKMDHKRRGALLFENNERFWMLAPERRGNADNDPTREESLGEYK 65
Qy 226 SR-VFDEKTYATRNSSRGKCLITNNEH--E-QMPTNGRKADKDNLTNFRMGTYI 25
Db 70 GENDRAEELLKIKHEVST--SSHVDACELCYVLSHGEGNHIVAYDAK-IEIQTULGEK 125
Qy 282 CKDNLTGRMLTITRDFAKESHGSDSA-LL-VILSHGEENYIIGVDPIETHEIYDLN 135
Db 128 GDKQCSITGVGPKFIETIACGSHOHV--P-LVPLDYV-D--HQT-DKLPD--N-VNQY- 175
Qy 340 AANNPRLANKPKIYEVQACGERGRDNGFYLDSDVCPALFIRGMDNRDGLFNFLGCLR 305
Db 176 -DAASVY-TLPAGADFLMCYSVAEGYYSHRETVNGSWYIOLDEMLARHGSLEFEFLT 235
Qy 400 PQVOOVRRKKPSQADILIRATTAQYVSNRNSARGSWFIOACVEYFTHAKMDVYELL 455
Db 234 LVNRRYSORAYDFECKDGAIGKKOYVCPFSMLTKLHF 271
Qy 460 EYNNKVA---CGFOTSGSNTLKQMEPMSTRLLKPFYF 494

```

RESULT	6	
ID	088550	PRELIMINARY;
AC	088550;	PRT; 303 AA

DT 01-NOV-1998 (TREMBLREL, 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL, 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL, 08, LAST ANNOTATION UPDATE)
 DE CASPASE-7
 OS RATTUS NORVEGICUS (RAT).
 OC EUARCTOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIROGNATHI; MORIDAE; MORINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SPLEEN;
 RA FORGHANI F., ROY S.;
 RT "Rat caspase-7 sequence."
 RU SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL, AF072124; G3243276; -;
 DR PROSITE; PS01121; CASPASE_HIS: 1.
 DR PROSITE; PS01122; CASPASE_CYS: 1.
 SQ SEQUENCE 303 AA; 34324 MW; BA732E7C CRC32;
 Query Match 9.1%; Score 328; DB 11; Length 303;
 Best Local Similarity 36.5%; Pred. No. 4,93e-44;
 Matches 99; Conservative 62; Mismatches 72; Indels 38; Gaps 20;

```

Db      59  IYLMDFEER-MKCKIILINKKNDKATGDVNRGTDKDALEKCFERSLGEFEVTVNDSCA 117
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      234 MYR-NFSSPRMCLLINNHEEQ---MPTNRGTAKDRDNLTNLFRMGKGYTV-ICKDLTG 288

Db      118 K-MODLLER-ASEDSHSACFACVLLSHGEBNLIYKQDVT-PINKULTAHFGDRCKTL 174
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      289 RCMLLTITDFDKHSHHGSDA-IL-VILSHGEBNVIYGVDDIPSTHETIYDLNANNAAPRL 346

Db      175 LEKPLFEITQACRGK-E-----LD--DGIDA-----DS--GPIND-TD-ANPR-----Y 211
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      347 ANKRPVEIVQACRGRRNRNGFPVLDSVQGYAPFLRGMDNRDGLFNFLOCVPRQOQVM 406

Db      212 -KIPYADFLFAYSTVPQYIYSRRNPGKSNFVQALCSITLNEHGKDEIFMQLITRVANDRA 270
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      407 RKKPSQADILIRYATATAYVSMRSARSSEWFIQAVCEVSTHAKMDVVELTEVNKKA 466

Db      271 RHFEESODDPRENEKQICPCVWSMLTKELY 301
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy      467 CGFQT-SQGS--NILKQMPKETSRLKKFEY 494

```

RESULT	7	
ID	042284	PRELIMINARY; PRT; 283 AA

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CASPASE-1.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: ARCHOSAURIA: AVES
OC NEOGNATHAE: GALLIFORMES: PHASIANTIDAE: PHASIANTINAE: GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN.
RA JOHNSON A.L., BRIDGHAM J.T., MUNKS M.W.:
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBD. DATA BANKS.
DR EMBL, AF031351, G2642241; -
DR PFAM: PF00655: ICE_P10: 1.
DR PFAM: PF00656: ICE_P20: 1.
SQ SEQUENCE 283 AA; 32622 MW; 7A72DE81 CRC32;

Query Match	8.3%;	Score 298;	DB 13;	Length 283;
Best Local Similarity	31.9%;	Pred. NO. 1.35e-37;		
Matches	82;	Conservative	56;	Mismatches 93;
			Indels	26;
			Gaps	19;

```
Db      45 LLLCNCRTKRNLRQRCGAENVDCYKMTKLLEGIGVYNVECHEDKTSQDENTVMKKFADHHDHL 104
        | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy     246 LIINHEHCQMPTRGTCRADKNLNLNPRCMGYTICDNLTGGMALTTINDFAKHESH- 304
        | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     105 TDSDFEIVMSMGSTGJCGTKSNCTDILSPNTIYENFNKHCORALMGRPKVYIOCCR 164
        |||:||||| || | : | : | : | : | : | : | : | : | : | : | : |
```

Db 61 ORRGVAFDPAFYDALRSTGHEGLAEVLEPLARVSDNAVEECPMSPASHRRSRALSPAG 120
 62 QRRGVAFDPAFYDALRSTGHEGLAEVLEPLARVSDNAVEECPMSPASHRRSRALSPAG 121
 Db 121 YTSPTVRHDSVSVSFTSYDITYSRARSRSRALHSSDRHNTSPPVNAFSPQSSA 180
 122 YTSPTVRHDSVSVSFTSYDITYSRARSRSRALHSSDRHNTSPPVNAFSPQSSA 181
 Db 181 NSFTGCSLGVSSSRNRSFKASGPTQYIHEEDMNPVADATISRYDEKTMNFESSP 240
 182 NSFTGCSLGVSSSRNRSFKASGPTQYIHEEDMNPVADATISRYDEKTMNFESSP 241
 Db 241 RGMCLINNEHEEOMPTNGTRKADKNTLNFRCMGYTVICKDNLTRGMLTTIRDFAKH 300
 242 RGMCLINNEHEEOMPTNGTRKADKNTLNFRCMGYTVICKDNLTRGMLTTIRDFAKH 301
 Db 301 ESHGSAIIVLISHGEEVNIIGVDIPITHEITLIDLANAANPRLANKPIYVQACSGE 360
 302 ESHGSAIIVLISHGEEVNIIGVDIPITHEITLIDLANAANPRLANKPIYVQACSGE 361
 Db 361 RRDNGEPVLDSDVGVPAFLRRGMDNRDGLFNLGCVRPQVOQVWRKKSQADILIAVAT 420
 362 RRDNGEPVLDSDVGVPAFLRRGMDNRDGLFNLGCVRPQVOQVWRKKSQADILIAVAT 421
 Db 421 TAOYVSWRSARGSWFIQACEVSTHAKDMDVLELVNKVAVDIPPAQ 472
 422 TAOYVSWRSARGSWFIQACEVSTHAKDMDVLELVNKVAVDIPPAQ 473

RESULT 2
 ID 093415 PRELIMINARY: PRT: 304 AA.
 AC 093415;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CASPASE-6.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA JOHNSON A.L., BRIDGEMAN J.T.;
 RT "Characterization of the chicken caspase-6 cDNA."
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF082329; G3435272; -.
 SQ SEQUENCE 304 AA; 34534 MW; D01C5D1B CRC32;

Query Match 10.4%; Score 375; DB 13; Length 304;
 Best Local Similarity 32.1%; Pred. No. 2,57e-54;
 Matches 94; Conservative 72; Mismatches 104; Indels 23; Gaps 16;

Db 28 NONTTEVDAEDKRRTPDAQOYKMHQRGVALIFNEHEFWMHLRLPDRGTLADRNNLK 87
 29 NONTTEVDAEDKRRTPDAQOYKMHQRGVALIFNEHEFWMHLRLPDRGTLADRNNLK 88
 Db 214 EEDNPFVDAPIISVFEDEKTYRNFSRPMCLINNEHE--E-OMPTNGTRKADKNTL 270
 215 EEDNPFVDAPIISVFEDEKTYRNFSRPMCLINNEHE--E-OMPTNGTRKADKNTL 271
 Db 88 RSLDLDGEVRIIPDLDKAEVLYKVEASRDYSNACFCVFLSGENDHYAYDA-QI 146
 89 RSLDLDGEVRIIPDLDKAEVLYKVEASRDYSNACFCVFLSGENDHYAYDA-QI 147
 Db 271 NLFRCMGYTVICKDNLTRGMLTTIRDFAKH-SHGDSALIVLISHGEEVNIIGVDIP 329
 272 NLFRCMGYTVICKDNLTRGMLTTIRDFAKH-SHGDSALIVLISHGEEVNIIGVDIP 330
 Db 147 KIETITNFRGDKQSLVGKPKIFIIIOACRGDKHD--PYLVO-DSYDS--K--D--ET 196
 148 KIETITNFRGDKQSLVGKPKIFIIIOACRGDKHD--PYLVO-DSYDS--K--D--ET 197
 Db 330 STEHETIDLANAANPRLANKPIYVQACRGERNDFPVLSDVGVPAFLRRGMDNRD 389
 331 STEHETIDLANAANPRLANKPIYVQACRGERNDFPVLSDVGVPAFLRRGMDNRD 390
 Db 197 TV-NOTE-V--DAAGV-TLPAGADFTMCSVAGQYSHETVNGSYIIDDCEALGKHG 251
 198 TV-NOTE-V--DAAGV-TLPAGADFTMCSVAGQYSHETVNGSYIIDDCEALGKHG 252
 Db 390 PLFNLCGCVRPQVOQVWRKKSQADILIAVATTAQYVSWRSARGSWFIQACEVSTHA 449
 391 PLFNLCGCVRPQVOQVWRKKSQADILIAVATTAQYVSWRSARGSWFIQACEVSTHA 450
 Db 252 SLIETETELLTVANRRVSHRKVDICRDINAIGKQIPCFASMLTKLFLYHPSK 304
 253 SLIETETELLTVANRRVSHRKVDICRDINAIGKQIPCFASMLTKLFLYHPSK 305
 Db 450 KDMVVELLTVENKRVAC-GEOTSGSN-I-LKQMPETSLKLLKFFWEAR 499
 451 KDMVVELLTVENKRVAC-GEOTSGSN-I-LKQMPETSLKLLKFFWEAR 500

RESULT 3

ID 018203 PRELIMINARY: PRT: 642 AA.
 AC 018203;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE Y48EB.13 PROTEIN.
 GN Y48EB.13.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MCMURRAY A.,
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
 RA CAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAURELLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK U., WILKINSON-SPROAT J., WOLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL NATURE 368:32-38(1994).
 DR EMBL: Z93393; E1354638; -.
 SQ SEQUENCE 642 AA; 73590 MW; B351D5D0 CRC32;

Query Match 9.9%; Score 356; DB 5; Length 642;
 Best Local Similarity 47.2%; Pred. No. 3.92e-50;
 Matches 60; Conservative 23; Mismatches 40; Indels 4; Gaps 4;

Db 363 YEMNSPRGTVLLISNENFKMERVGTQDEVTLYLFOKLYTVICKRNLAESEMLTA 422
 364 YEMNSPRGTVLLISNENFKMERVGTQDEVTLYLFOKLYTVICKRNLAESEMLTA 423
 Db 235 YRNFSRPMCLINNEHEFDMPTNTRKADKNTLNFRCMGYTVICKDNLTRGMLTT 294
 236 YRNFSRPMCLINNEHEFDMPTNTRKADKNTLNFRCMGYTVICKDNLTRGMLTT 295
 Db 423 IKERA-EMAHDSIILFLSHDGAGSVFGIDMPVAVMEVSTYL-AYH-QNLLAKPKV 479
 424 IKERA-EMAHDSIILFLSHDGAGSVFGIDMPVAVMEVSTYL-AYH-QNLLAKPKV 480
 Db 295 IRDFAKHSHSDSALIVLISHG--ENVIGVDIPITHEITLIDLANAANPRLANKPIY 353
 296 IRDFAKHSHSDSALIVLISHG--ENVIGVDIPITHEITLIDLANAANPRLANKPIY 354
 Db 480 AVSACRG 486
 481 AVSACRG 487
 Db 354 FVQACRG 360
 355 FVQACRG 361

RESULT 4
 ID 055194 PRELIMINARY: PRT: 452 AA.
 AC 055194;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE NEDD2/ICH-1.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98087427.
 RA SATO N., MILLIGAN C.E., UCHIYAMA Y., OPPENHEIM R.W.;
 RT "Cloning and expression of the cDNA encoding rat caspase-2."
 RL GENE 202:127-132(1997).
 DR EMBL: U77933; G2169706; -.
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS01122; CASPASE_CYS; 1.
 SQ SEQUENCE 452 AA; 50728 MW; 0FA96CD3 CRC32;

Query Match 9.8%; Score 354; DB 11; Length 452;
 Best Local Similarity 33.8%; Pred. No. 1.07e-45;

M P E R E H
(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:39:02 1999; Msrch time 31.00 Seconds

Tabular output not generated. 885.516 Million cell updates/sec

Title: >US-08-287-669-19

Description: (1-503) from US08287669.pep

Sequence: 1 MMRDRSLRLNIMFSSH.....MTSRLKVFWEPEARNSAV 503

Scoring table: PAM 150

Gap 11

Searched: 179066 segs, 54579741 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database:

sptrembl9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mmc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.potent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 49.838; Variance 85.439; scale 0.583

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	3307	91.8	495	5	002229	C48D1.2 PROTEIN.	0.00e+00
2	375	10.4	304	13	093415	CASPASE-6.	2.57e-54
3	356	9.9	642	5	018203	Y48E1B.13 PROTEIN.	3.92e-50
4	354	9.8	452	11	055194	NEDD2/ICH-1.	1.07e-49
5	350	9.7	277	11	035397	CASPASE 6.	8.07e-49
6	328	9.1	303	11	088550	CASPASE-7.	4.93e-44
7	288	8.3	283	13	042284	CASPASE-1.	1.35e-37
8	285	8.2	283	13	093417	CASPASE-3.	5.87e-37
9	269	7.5	182	6	077623	CASPASE-7.	1.73e-31
10	268	7.4	377	4	075601	EVOLUTIONARILY RELATED	2.80e-31
11	257	7.1	257	11	089094	CASPASE-14.	5.33e-29
12	259	6.1	480	11	089110	CASPASE-8.	2.63e-21
13	207	5.7	422	11	035669	FLICE PARTIAL CDS (FLI	6.03e-19
14	206	5.7	464	4	014676	APOPTOTIC CASPASE MCIS	9.44e-19
15	202	5.6	366	5	002433	DREDD.	5.65e-18
16	202	5.6	488	5	076797	DREDD ISOFORM GAMMA.	5.65e-18
17	202	5.6	488	5	076797	DREDD ISOFORM DELTA.	5.65e-18
18	142	3.9	1238	11	088879	APOPTOTIC PROTEASE ACT	5.18e-07
19	122	3.4	470	5	022919	COSMID C37C3.	9.25e-04
20	114	3.2	445	4	014673	FLAME-1.	1.53e-02

21	114	3.2	449	4	043618	I-FLICE ISOFORM 3.	1.53e-02
22	114	3.2	462	4	060458	USURPIN-BETA.	1.53e-02
23	114	3.2	480	4	015356	CASH ALPHA PROTEIN.	1.53e-02
24	114	3.2	480	4	015519	CASPR.	1.53e-02
25	112	3.1	210	14	066677	ORF E10.	3.03e-02
26	111	3.1	323	11	088602	GAMMA-2 SUBUNIT OF VOL	4.24e-02
27	113	3.1	348	4	043620	I-FLICE ISOFORM 5.	2.15e-02
28	110	3.1	433	2	052401	THIOESTERASE.	5.93e-02
29	112	3.1	442	4	043619	I-FLICE ISOFORM 4.	3.03e-02
30	112	3.1	480	4	015137	CASPASE-LIKE APOPTOSIS	3.03e-02
31	107	3.0	744	10	065375	F12F1.9 PROTEIN.	1.60e-01
32	103	2.9	423	5	P90904	KOZB12.7 PROTEIN.	5.84e-01
33	103	2.9	495	5	052406	PUTATIVE 54.5 KDA PROT	2.22e-01
34	106	2.9	703	6	P79122	PINTN.	2.22e-01
35	104	2.9	1095	4	060300	KIA00533 PROTEIN (FRAG	4.22e-01
36	106	2.9	1288	4	015038	KIA00324 (FRAGMENT).	2.22e-01
37	106	2.9	1791	4	060382	KIA00324 (FRAGMENT).	2.22e-01
38	101	2.8	328	2	050356	TRANSPASASE.	1.10e+00
39	101	2.8	330	10	043041	CHALCONE SYNTHASE (EC	1.10e+00
40	102	2.8	358	4	076004	ATP(GTP)-BINDING PROTE	8.02e-01
41	101	2.8	424	5	017469	B0284.2 PROTEIN.	1.10e+00
42	101	2.8	452	10	023146	HNRNP-LIKE PROTEIN.	1.10e+00
43	102	2.8	1645	5	099140	BETA-H SPECTRIN (FRAGM	8.02e-01
44	101	2.8	2761	5	018447	CODED FOR BY C. ELEGAN	1.10e+00
45	101	2.8	2810	5	020456	HUM-4 PROTEIN.	1.10e+00

ALIGNMENTS

RESULT 1
ID 002229 PRELIMINARY; PRT; 495 AA.

AC 002229;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DI 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE C48D1.2 PROTEIN.

OS CAENORABDITIS ELEGANS.

OC EUDARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORABDITIS.

RA [1]
PP SEQUENCE FROM N.A.

RA BURTON J.;

RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]
RP SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R.; AINSOUGH R.; ANDERSON K.; BAYNES C.; BERKS M.;

RA BOWFIELD J.; BURTON J.; CONNELL M.; COPSEY T.; COOPER J.; COULSON A.;

RA CRAXTON M.; DEAR S.; DU Z.; DURBIN R.; FAVELLO A.; FULTON L.;

RA GARDNER A.; GREEN P.; HAWKINS T.; HILLIER L.; JER M.; JOHNSTON L.;

RA JONES M.; KERSHAW J.; KIRSTEN J.; LAISTER N.; LATREILLE P.;

RA LIGHTNING J.; LLOYD C.; MCURRAY A.; MORTIMORE B.; O'CALLAGHAN M.;

RA PARSONS J.; PERCY C.; RIKEN T.; ROOPER A.; SAUNDERS D.; SHOMKKEEN R.;

RA SMALDON N.; SMITH A.; SONNHAMMER E.; STADEN R.; SULSTON J.;

RA THIERRY-WIEG J.; THOMAS K.; VAUDIN M.; VAUGHAN K.; WATERSTON R.;

RA WATSON A.; WEINSTOCK L.; WILKINSON-SPROAT J.; WOHLIDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RL NATURE 368:32-38(1994).

DR EMBL; Z81048; E1344743; .

DR PROSITE; PS01121; CASPASE_HIS; 1.

DR PROSITE; PS01122; CASPASE_CYS; 1.

SO SEQUENCE 495 AA; 55549 MW; 7DC9ED9D CRC32;

Query Match 91.8%; Score 3307; DB 5; Length 495;

Best Local Similarity 98.5%; Pred. No. 0.00e+00;

Matches 465; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 1 MMRDRSLRLNIMFSSHLKVDLEIIVLIAKOVLSNDGDMNSGTVRKRREIVKAV 60
QY 2 MMRDRSLRLNIMFSSHLKVDLEIIVLIAKOVLSNDGDMNSGTVRKRREIVKAV 61

RT and CED-3.";
 RL ONCOGENE 13:749-755(1996).
 RN [2]
 RP SEQUENCE OF 30-241 FROM N.A.
 RC TISSUE-OVARY;
 RX MEDLINE: 96042508.
 RA FLAWS J.A., KUGU K., TROBICH A.M., DESANTI A., TILLY K.I.,
 RA HIRSHFIELD A.N., TILLY J.L.,
 RT "Interleukin-1 beta-converting enzyme-related proteases (IRPs) and
 RT mammalian cell death: dissociation of IRP-induced oligonucleosomal
 RT endonuclease activity from morphological apoptosis in granulosa cells
 RT of the ovarian follicle."
 RL ENDOCRINOLOGY 136:5042-5053(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 97184204.
 RA NI B., WU X., DU Y., SU Y., HAMILTON-BYRD E., ROCKEY P.K.,
 RA ROSTECK P. JR., POIRIER G.G., PAUL S.M.,
 RT "Cloning and expression of a rat brain interleukin-1beta-converting
 RT enzyme (ICE)-related protease (IRP) and its possible role in
 RT apoptosis of cultured cerebellar granule neurons."
 RL J. NEUROSCI. 17:1561-1569(1997).
 RN [4]
 RP SEQUENCE OF 1-264 FROM N.A.
 RA YAKOVLEV A.G.;
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION. AT THE ONSET OF APOPTOSIS IT
 CC PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A
 CC 216-ASP-1-GLY-217 BOND. CLEAVES AND ACTIVATES STEROL REGULATORY
 CC ELEMENT BINDING PROTEINS (SREBPS) BETWEEN THE BASIC HELIX-LOOP-
 CC HELIX LEUCINE ZIPPER DOMAIN AND THE MEMBRANE ATTACHMENT DOMAIN.
 CC CLEAVES AND ACTIVATES CASPASE-6, -7 AND -9 (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNIT
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LIVER, AND MUSCLE
 CC BUT NOT IN KIDNEY OR TESTIS.
 CC -1- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN NEURON-ENRICHED REGIONS
 CC OF THE DEVELOPING BRAIN, BUT DOWN-REGULATED TO LOW LEVELS IN THE
 CC ADULT BRAIN.
 CC -1- PTM: CLEAVAGE BY GRANTZYME B, CASPASE-6, -8 AND -10 GENERATES THE
 CC TWO ACTIVE SUBUNITS. ADDITIONAL PROCESSING OF THE ACTIVATED
 CC PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
 CC CASPASE-7 PROTEASE AND THE LARGE SUBUNIT OF CPP32 ALSO OCCUR AND
 CC VICE VERSA (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
 CC CASPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U49930; G1518936; -
 DR EMBL: U34685; G1004371; -
 DR EMBL: U84410; G1814027; -
 DR EMBL: U58656; G1381643; -
 DR PROSITE: PS01121; CASPASE_HIS_1.
 DR PROSITE: PS01122; CASPASE_CYS_1.
 DR PFAM: PF00655; ICE_P10; 1.
 DR PFAM: PF00656; ICE_P20; 1.
 DR HSSP: P42574; 1PAU.
 KW HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
 FT PROPEP 1 9
 FT PROPEP 10 28 BY SIMILARITY.
 FT CHAIN 29 175 APOPAIN P17 SUBUNIT.
 FT CHAIN 176 277 APOPAIN P12 SUBUNIT.

FT	ACT_SITE	121	121	BY SIMILARITY.
FT	ACT_SITE	163	163	BY SIMILARITY.
FT	CONFLICT	25	29	KSMDS -> OVD (IN REF. 4).
FT	CONFLICT	170	170	C -> S (IN REF. 2).
FT	CONFLICT	178	178	T -> A (IN REF. 2).
FT	CONFLICT	182	182	M -> V (IN REF. 2).
FT	CONFLICT	187	187	I -> K (IN REF. 2).
FT	CONFLICT	190	190	E -> G (IN REF. 3).
FT	CONFLICT	199	199	T -> S (IN REF. 2).
FT	CONFLICT	211	211	D -> G (IN REF. 2).
FT	CONFLICT	236	236	L -> I (IN REF. 4).
FT	CONFLICT	245	245	T -> M (IN REF. 3).
SO	SEQUENCE	277 AA;	31491 MW;	0E106140 CRC32;

Query Match 7.9%; Score 284; DB 1; Length 277;
 Best Local Similarity 33.8%; Pred. No. 1,86e-34;
 Matches 91; Conservative 63; Mismatches 76; Indels 39; Gaps 20;

Db	37	YKMDPE-MGICITLNKNEHKTGMSARNGTDVDAANLRETFMALKEVRRNNDLT-RE 94
Qy	235	YR-NFSSFRGMCLTIINHEFEQ--MPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRG 290
Db	95	EIMELMDSVSKEDSKRSFVCVLTLSHGDEGVIFGTNG-PVDLKLTSFFRGDYCSLTG 153
Qy	291	MULTIRDFAKHESHGD-SALT-VILSHGEENVIIIGVDIPISITHEIYDLNANAPRLAN 348
Db	154	KPKLFIIOACRGK--E-----LDC--GIET--DSG--TD-D-----D-MAC-----OKI--- 187
Qy	349	KPKLVFOACGERNDGFVLDVSYCPAFILRGMDNRGPFNPLGCVRPVQVQWKR 408
Db	188	-PVADLVAYSTAPGYISWRNSRDSGWFQSLQCAMIKLYAHLEFMHILTRVNRVATE 246
Qy	409	KPSQADILIRATYTAQYVSRNSARGSGWFIQAVCEVFSTHAKMDVVELLTVNKKVACG 468
Db	247	FESFSLDAPFAKQOIPCIYSMLTKELYE 275
Qy	469	FQT-S-QGS-NILKQEMTSRLKKFYF 494

Search completed: Tue Aug 10 11:38:44 1999
 Job time : 59 secs.

[illegible]

DR	EMBL, U04332; G1245144; -			
DR	PROSITE; PS01121; CASPASE_HIS; 1.			
DR	PROSITE; PS01122; CASPASE_CYS; 1.			
DR	PFAM; PF00655; ICE_P10; 1.			
DR	PFAM; PF00656; ICE_P20; 1.			
DR	HSPD; P42574; 1PAU.			
KM	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.			
FT	PROPEP	1	23	
FT	CHAIN	24	198	
FT	PROPEP	199	206	CASPASE-7 SUBUNIT P20.
FT	CHAIN	207	303	BY SIMILARITY.
FT	ACT_SITE	144	144	CASPASE-7 SUBUNIT P11.
FT	ACT_SITE	186	186	BY SIMILARITY.
SO	SEQUENCE	303 AA;	34037 MW;	5C119AD4 CRC32;
Query Match				
Best Local	Similarity	9.1%;	Score 328;	DB 1; Length 303;
Matches	99; Conservative	36.7%;	Pred. NO. 7.74e-44;	
		59;	Mismatches 76;	Indels 36; Gaps 18.

[illegible]

```

Db      272 HFE5QCDDPCFENKROICPCWMLKELEYE 301
        |:::| |::| |::| |::| |::|
Oy      468 GFOT-SQGS--NLIKOMPDMISRLUKRYE 494

RESULT 10
ID ICE7_MOUSE STANDARD: PRT: 303 AA.
AC P97864; 008669;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE CASPASE-7 PRECURSOR (EC 3.4.22.-) (LICE2 CYSTEINE PROTEASE) (APOPTOTIC
DE PROTEASE MOH-3).
DN CASP7 OR MOH3 OR LICE2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MORIDAE; MORINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELTAL MUSCLE;
RX MEDLINE: 97224489.
RA JOAN T.S.-C., MCNEEHE I.K., ARGENTO J.M., JENKINS N.A., GILBERT D.J.,
RA COLEMAN N.G., FLETCHER F.A.;
RT "Identification and mapping of Casp7, a cysteine protease resembling
RL CPB32 beta. Interleukin-1 beta converting enzyme, and CED-3.";
RN GENOMICS 40:86-93(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97236307.
RA MURASA T., KHOROKU Y., TSUKAHARA T., MOMOI M.Y., KIMURA I.,
RA MOMOI T.;
RT "wortmannin enhances CPP32-like activity during neuronal
RT differentiation of P19 embryonal carcinoma cells induced by retinoic
RL acid.";
RN BIOCHEMA. BIOPHYS. RES. COMMUN. 232:192-197(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/AN.
RX MEDLINE: 97190206.
RA VAN DE CRAEN M., VANDENABEELE P., DECLERCQ W., VAN DEN BRANDE I.,
RA VAN LOO G., MOLEMANS F., SCHOTTE P., VAN CRIEKINGE W., BEYVERET R.,
RA FIERS W.;
RT "Characterization of seven murine caspase family members.";
RL FEBS LETT. 403:61-69(1997).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES AND ACTIVATES STERO
CC REGULATORY ELEMENT BINDING PROTEINS (SREBPS). OVEREXPRESSION
CC PROMOTES PROGRAMMED CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF A 20 KD (P20) AND A 11 KD (P11) SUBUNIT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, LUNG, LIVER, AND
CC KIDNEY. LOW LEVELS IN SPLEEN, SKELETAL MUSCLE, AND TESTIS. NO
CC EXPRESSION IN THE BRAIN.
CC -1- PM: CLEAVAGES BY GRANZYME B OR CASPASE-10 GENERATE THE TWO ACTIVE
CC SUBUNITS. PROPEPTIDE DOMAINS CAN ALSO BE CLEAVED EFFICIENTLY BY
CC CBP32 PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC CASPASE-7 AND THE LARGE SUBUNIT OF CPP32 PROTEASE ALSO OCCUR AND
CC VICE VERSA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-ch).
CC -----
DR EMBL: U67321; G1894917; ALU_INIT.
DR EMBL: D86353; D1020510; -.
DR EMBL: Y13086; E315506; -.
DR MGJ: MG1109383; CASP7.
```

DE (ICE-LAP3) (APOPTOTIC PROTEASE MCH-3) (SRBP CLEAVAGE ACTIVITY 2)
DE (SCA-2).

DR	PROSTE:PPS01122; CASPASE_CYS; 1.
DR	PFAM; PF00619; CARD; 1.
DR	PFAM; PF00655; ICE_P10; 1.
DR	PFAM; PF00656; ICE_P20; 1.
DR	HSP; P29466; IIBC.
KW	HYDROLASE; THIOL PROTEASE; APOPTOSIS; ZMOGEN.
FT	PROPEP 1 ? POTENTIAL.
FT	CHAIN 7 316 CASPASE-2 SUBUNIT 1 (BY SIMILARITY).
FT	CHAIN 317 435 CASPASE-2 SUBUNIT 2 (BY SIMILARITY).
FT	AOL_SITE 260 260 BY SIMILARITY.
FT	AOL_SITE 303 303 BY SIMILARITY.
FT	MUTAGEN 303 303 C->G: LOSS OF FUNCTION.
FT	CONFLICT 71 78 MISSING (IN REF. 1).
SO	SEQUENCE 435 AA; 48896 MW; 5461BFE CRC32;
Query Match 10.0%; Score 359; DB 1; Length 435;	
Best Local Similarity 33.5%; Pred. NO. 1.39e-50;	
Matches 90; Conservative 59; Mismatches 100; Indels 20; Gaps 15,	
Dd	175 YRLSQPGMLVLVSNVFETGEKDLDFPRSGGDVDHTLLVTFLKLGVNVAHLHQTOEM 234 :: :
Oy	235 YNESPSRGMCLLIINEHF--E-QPIIRNGAKARKDNLTINFRCMGTYVICDKDLGRM 291 :: :
Dd	235 QEKLNFAQLPALHRYTDCVAVALLSHGEVGIVGDKLLOLFVEFRPLFNANCPSLONK 294 :: :
Oy	292 LLTIIDFPAKHESH-GDSAILVLISHGENYIICVDIPISTHEIIDLLMAANAPLANK 349 :: :
Dd	295 PKMFIOACRDENDRG--V-DODDG-KNH-TQS-PCEES--D-AG--KEELMK-RL- 341 :: :
Oy	350 PRIVYGVCCKRRNRNGPVLDSDVPALFRMRNDRPLENFGLCVBPVOOVARRKK 409 :: :
Dd	342 PTRSDMIGVALCNGNAAMRTKRGSWTIEALTGFSEPARCDMHVADLVAKNALIKERE 401 :: :
Oy	410 PSQAIDLIRATTAQQYWSRNSSANGSFIAQVCFSTHAKDMDVVELLEVKKV-AC- 467 :: :
Dd	402 GYAPOTEFHRCKEMSFCITLCOOLYFP 430 :: :
Oy	468 GFOTSOGSNIILKOMEMTSRLKKFYWP 496 :: :
RESULT 7 STANDARD; PRT; 276 AA.	
ID	ICE6.MOUSE
AC	008738:
Dt	01-NOV-1997 (REL. 35, CREATED)
Dt	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
Dt	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE	CASPASE-6 PRECURSOR (EC 3.4.22.-) (APOPTOTIC PROTEASE MCH-2).
GN	CASP6.
OS	MUS MUSCULUS (MOUSE).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC	RODENTIA; SCUROGNATHI; MORIDAE; MORINAE; MUS.
RN	[1]
RP	SEQUENCE FROM N.A.
RP	STRAIN=C3H/AN;
RC	MEDLINE: 97190206.
RA	VAN DE CAEN W., VANDENAEBELE P., DECIERO W., VAN DEN BRANDE I., VAN LOO G., MOLEMANS F., SCHOTTE P., VAN CREKINGE W., BEYAERT R., FIERS W.;
RT	"Characterization of seven murine caspase family members.";
RL	FELS LETT. 403:61-69(1997).
CC	-1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES RESPONSIBLE FOR APOPTOSIS EXECUTION. CLAVES POLI(ADP-RIBOSE) POLYMERASE IN VITRO, AS WELL AS LAMINS. OVEREXPRESSION PROMOTES PROGRAMMED CELL DEATH (BY SIMILARITY).
CC	-1- SUBUNIT: HETERODIMER OF A 18 KD (P18) AND A 11 KD (P11) SUBUNIT (BY SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC	-1- PTM: CLEAVAGES BY CPP32, CASPASE-8 OR -10 GENERATE THE TWO ACTIVE SUBUNITS (BY SIMILARITY).
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, LIVER, KIDNEY, TESTIS, AND HEART. LOWER LEVELS IN SPLEN, SKELETAL MUSCLE, AND BRAIN.
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE

QY 46 / C-GHQTSQGSNLLKQMPENISRLKKFYFWP 496

[illegible]

Query Match	7.7%; Score 278; DB 2; Length 402,
-------------	------------------------------------

Search completed: Tue Aug 10 11:41:53 1999
Job time : 58 secs.

QY	356	QACRGERRNG	366	:
RESULT	7			
ENTRY		C56084	#type complete	
TITLE		interleukin- β converting enzyme delta isozyme - human		
ORGANISM		#formal_name Homo sapiens #common_name man		
DATE		03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 17-Mar-1999		
ACCESSIONS		C56084		
REFERENCE		A56084		
#authors		Alnemri, E.S.; Fernandes-Alnemri, T.; Litwack, G.		
#journal		J. Biol. Chem. (1995) 270:4312-4317		
#title		Cloning and expression of four novel isoforms of human interleukin- β converting enzyme with different apoptotic activities.		
#cross-references		MUID:95181414		
#accession		C56084		
#status		preliminary		
#molecule_type		mRNA		
#residues		1-263	#label ALN	
#cross-references		GB:U13699; NID:g717043; PID:g717044		
GENETICS				
#gene		IL13CE		
KEYWORDS		alternative splicing		
SUMMARY		length 263	#molecular-weight 29821	#checksum 617
Query Match		8.0%	Score 289;	DB 2; Length 263;
Best Local Similarity		35.4%	Pred. No. 4,45e-30;	
Matches		56;	Conservative 38;	Mismatches 53; Indels 11; Gaps 7
Db	50	RIMWOKSAEIVPIMDKSSRTLRALITICNEEEDSIPRTGAEDVITGTMQLQNLGYSVDV	109	
QY	227	RVFDEKT--MYRNFS-SPRG-MCLINNEHFQMPTRNCTKADKNLNLFRCMGYVIC	282	
Db	110	KNLTASDMTELEAPAPRPHKHTSDSTFLVEMSHGIRGICGKHSQVDPDILQNAIF	169	
QY	283	KDNLTRKMLITTFIDFAKHESH--GDSALIVLISHGEENVIIIG---VDDIP-I-STHEIY	335	
Db	170	NMLNTKNCPSLKDPRKVIILIQACRGDNVSWRHPITMSV	207	
QY	336	DLNANANPRLANKPIVFOACRGERRNGPFVLDV	373	
RESULT	8			
ENTRY		I39005	#type complete	
TITLE		cysteine proteinase (EC 3.4.22.-) Yama - human		
ALTERNATE_NAMES		cysteine proteinase CPP32		
ORGANISM		#formal_name Homo sapiens #common_name man		
DATE		16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 16-Feb-1996		
ACCESSIONS		I39005		
REFERENCE		A56924		
#authors		Tewari, M.; Quan, L.T.; O'Rourke, K.; Desnoyers, S.; Zeng, V.M.		
#journal		Cell (1995) 81:801-809		
#title		Yama/ CPP32 beta, a mammalian homolog of CED-3, is a CrmA-inhibitable protease that cleaves the death substrate poly(ADP-ribose) polymerase.		
#cross-references		MUID:95292347		
#accession		I39005		
#status		preliminary		
#molecule_type		mRNA		
#residues		1-277	#label RES	
#cross-references		EMBL:U6943; NID:9857568; PID:9857569		
KEYWORDS		apoptosis; cysteine proteinase; hydrolase		
SUMMARY		length 277	#molecular-weight 3108	#checksum 4536
Query Match		7.9%	Score 285;	DB 2; Length 277;
Best Local Similarity		33.2%	Pred. No. 2,34e-29;	
Matches		89;	Conservative 66;	Mismatches 76; Indels 37; Gaps 18;

```

QY 301 HESHGSAIIVLISHGEENVIGVDIPISTHEITDILNAANAPRLANKPIVFOACG 360
Db 361 ERSDNPFVLDVGVAPFLRGMNDRDLFNFGLGCVRPQVOYWRKKPSOADIILIA 420
QY 361 ERSDNPFVLDVGVAPFLRGMNDRDLFNFGLGCVRPQVOYWRKKPSOADIILIA 420
Db 421 TTAQVYVSWNSRNSAGSWFIOAVCEVSTHAKMDVVELLTVNKKVACGFGTSGSNILK 480
QY 421 TTAQVYVSWNSRNSAGSWFIOAVCEVSTHAKMDVVELLTVNKKVACGFGTSGSNILK 480
Db 481 MPMTSRLKKFYFMPPEARNSAV 503
QY 481 MPMTSRLKKFYFMPPEARNSAV 503

RESULT 2
ENTRY 2
TITLE G02635 #type complete
ORGANISM ICB-LAP6 - human
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
ACCESSIONS G02635
REFERENCE H01513
AUTHORS Duan, H.; Orth, K.; Chinaiyan, A.M.; Poirier, G.G.; Froelich, C.J.; He, W.W.; Dixit, V.M.
SUBMISSION submitted to the EMBL Data Library, April 1996
#accession G02635
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-416 #label DUA
#cross-references EMBL:U056390; NID:g1336026; PID:g1336027
SUMMARY #length 416 #molecular_weight 46322 #checksum 797

Query Match 10.3%; Score 372; DB 2; Length 416;
Best Local Similarity 32.2%; Pred. No. 2,08e-45;
Matches 77; Conservative 65; Mismatches 79; Indels 18; Gaps 14;

Db 159 PCGHCLINNVNFCRESGLTRFGSNIDCEKLRFRSSLFHVEVKGDLTAKKVALLE 218
QY 241 PRGMCCLINNEHF--EQ-MPTNRGTAKADKNDLTLFRMGVTVICDKNLTGRMGLTIRD 297
Db 219 IARODHAGALCCVAVILSHGCSASHLOFPAVYGTDCGPVSEKIVNINGTSCPSLGK 278
QY 298 FAHSHSG--DSAILVILSHG--EE-NV-I--I-GVDDIPISTHEITDILNAANAPRLANK 349
Db 279 PKLFFIOACGEGKDHGFV--AST-S--PEDESPG--SNPEPDATFOEGARTEDOLDAT-S 333
QY 350 PKIVFOACGGERDRNGFPVLDVGVAPFLRGMNDRDLFNFGLGCVRP--QVOQVWR 407
Db 334 SLPPSDIEVYSITFFPGFVSWRDPKSGSVYETLDTFEQMAHSEDLQSLILRVANAVS 392
QY 408 KKPQOADIILIRYATTAQVYVSWNSRNSAGSWFIOAVCEVSTHAKMDVVELLTVNKKVA 466

RESULT 3
ENTRY 3
TITLE A54821 #type complete
ORGANISM apoptotic regulator ICH-1, stimulatory form L - human
DATE 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 17-Mar-1999
ACCESSIONS A54821
REFERENCE A54821
AUTHORS Wang, L.; Miura, M.; Bergeron, L.; Zhu, H.; Yuan, J.
#journal Cell (1994) 78:739-750
#title Ich-1, an ICB-1-related gene, encodes both positive and negative regulators of programmed cell death.
#cross-references MUID:94373811
#accession A54821
#status preliminary
#molecule_type mRNA
#residues 1-435 #label WAN
#cross-references GB:U13021; NID:g537291; PID:g537292

```

```

KEYWORDS alternative splicing; apoptosis
SUMMARY #length 435 #molecular_weight 48855 #checksum 5880

Query Match 10.3%; Score 371; DB 2; Length 435;
Best Local Similarity 34.7%; Pred. No. 3.21e-45;
Matches 94; Conservative 58; Mismatches 95; Indels 24; Gaps 18;

Db 175 YRLOSOPRGALVLSNVHTEGKELEFRSGGDVDTLVLTKLGVNHYVLC--DO--TAO 232
QY 235 YRNSSPRGMLIINNEHF--E--QMPTRNGTKADKNDLTLFRMGVTV-I--CKDMLTGR 289
Db 233 EMOKLONFQOLPAHRYTDCIYALLSHGVEGATVYGDGKLLOLQVFOFQFNDANCPSLQ 292
QY 290 GMLITIRDFAKHESH--GDSAILVILSHGEENVIGVDIPISTHEITDILNAANAPRLA 347
Db 293 NKPRMFIQACRGDETRG--V--DOODG--KN--HAGSPGCEES--D--AG--KEKLPKM--R 340
QY 348 NKPRIVFOACRGERRNGFPVLDVGVAPFLRGMNDRDLFNFGLGCVRPQVOYWR 407
Db 341 L-PTRSDMICGYACLKTAAAMRTKRSWYIEALAOYFSERACDMHAYDMLVYNALIKD 399
QY 408 KKPQOADIILIRYATTAQVYVSWNSRNSAGSWFIOAVCEVSTHAKMDVVELLTVNKKV-A 466
Db 400 REGYAPGTEFHRCKEMSEYCTLCRHLVLP 430
QY 467 C-GFOTSGSNILKQMPMTSRLKKFYFMP 496

RESULT 4
ENTRY 4
TITLE JC6507 #type complete
ORGANISM caspase-2 - Rat
DATE 03-Sep-1998 #sequence_revision 03-Sep-1998 #text_change 03-Sep-1998
ACCESSIONS JC6507
REFERENCE JC6507
AUTHORS Sato, N.; Milligan, C.E.; Uchiyama, Y.; Oppenheim, R.W.
#journal Gene (1997) 202:127-132
#title Cloning and expression of the cDNA encoding rat caspase-2.
#cross-references MUID:98087427
#accession JC6507
#status preliminary
#residues 1-452 #label SAT
#cross-references GB:U77933
SUMMARY #length 452 #molecular_weight 50728 #checksum 6980

Query Match 9.8%; Score 354; DB 3; Length 452;
Best Local Similarity 33.8%; Pred. No. 4.96e-42;
Matches 91; Conservative 58; Mismatches 100; Indels 20; Gaps 15;

Db 192 YRLOSOPRGALVLSNVHTEGKELEFRSGGDVDTLVLTKLGVNHYVLDQTAQM 251
QY 235 YRNSSPRGMLIINNEHF--E--QMPTRNGTKADKNDLTLFRMGVTVICDKDLTGRM 291
Db 252 QEKLONFQOLPAHRYTDCIYALLSHGVEGATVYGDGKLLOLQVFOFQFNDANCPSLQK 311
QY 292 LITIRDFAKHESH--GDSAILVILSHGEENVITVDIPISTHEITDILNAANAPRLANK 349
Db 312 KPMFIQACRGDETRG--V--DOODG--KNHAG--S--PGCEES--D--AG--KEELKRM--RL- 358
QY 350 PKIVFOACGGERDRNGFPVLDVGVAPFLRGMNDRDLFNFGLGCVRPQVOYWRKK 409
Db 359 PTRSDMICGYACLKTAAAMRTKRSWYIEALQVFSERACDMHAYDMLVYNALIKERE 418
QY 410 PSQADILIRYATTAQVYVSWNSRNSAGSWFIOAVCEVSTHAKMDVVELLTVNKKV-AC- 467
Db 419 GYAPGTEFHRCKEMSEYCTLCQOVLVLP 447
QY 468 GFOTSGSNILKQMPMTSRLKKFYFMP 496

RESULT 5
ENTRY 5
TITLE B54821 #type complete

```

MORSE
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MORSE protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 10 11:40:55 1999; MasPar time 22.02 Seconds
Tabular output not generated. 915.459 Million cell updates/sec

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 3603
Sequence: 1 MMRDRSLRLERNIMFSSH.....MTSRLKKEFYFPEARNSAV 503

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r60
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 50.064; Variance 100.638; scale 0.497

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description	Pred. No.
No.	Score	%				
1	3591	99.7	503	2	A49429 interleukin-1 beta-co	0.00e+00
2	372	10.3	416	2	G02635 ICE-LAP6 - human	2.08e-45
3	371	10.3	435	2	A54821 apoptosis regulator I	3.21e-45
4	354	9.8	452	3	JC6507 caspase-2 - Rat	4.96e-42
5	307	8.5	312	2	B54821 apoptosis regulator I	2.41e-33
6	288	8.3	182	2	I67436 interleukin-1-beta-co	1.05e-31
7	289	8.0	263	2	C56084 interleukin-1-beta-co	4.45e-30
8	285	7.9	277	2	I39005 cysteine proteinase (5.35e-29
9	283	7.9	277	2	A55315 interleukin-1 beta co	4.22e-28
10	286	7.9	311	2	B56084 interleukin-1-beta con	1.54e-29
11	286	7.9	383	2	A56084 interleukin-1-beta con	1.54e-29
12	286	7.9	404	2	A42677 interleukin-1 beta co	1.54e-29
13	279	7.7	377	2	A57511 interleukin-1 beta co	2.79e-28
14	278	7.7	399	2	I48911 interleukin-1 beta co	4.22e-28
15	278	7.7	402	2	A46495 IL-1 beta convertase (4.22e-28
16	271	7.5	212	2	I67437 cysteine proteinase (7.50e-27
17	272	7.5	364	2	S62183 cysteine proteinase I	4.97e-27
18	272	7.5	401	2	A47258 interleukin-1 beta co	4.97e-27
19	272	7.5	418	2	B57511 interleukin-1 beta co	4.97e-27
20	267	7.4	277	2	JC5410 CPP32 protein - mouse	3.86e-26
21	253	7.0	277	2	S64710 cysteine proteinase (1.14e-23
22	148	4.1	136	2	I53300 interleukin-1-beta-co	1.78e-06
23	114	3.2	243	2	A42293 pyridoxal phosphate b	8.90e-02

ENTRY	1	ALIGNMENTS
24	115	3.2 376 2 S52137 MID2 protein - yeast
25	112	3.1 210 2 S55671 hypothetical protein
26	111	3.1 272 2 S59042 splicing factor SRP40
27	111	3.1 284 1 XYSOHA site-specific DNA-met
28	112	3.1 506 1 W2ML47 E2 protein - human pa
29	108	3.0 221 2 C64029 hypothetical protein
30	108	3.0 221 2 B42701 PR264 protein - chick
31	107	3.0 280 2 A53189 apoptosis suppressor
32	107	3.0 604 2 S68449 apoptosis inhibitor h
33	104	2.9 238 2 A57198 splicing factor, arg1
34	103	2.9 471 2 S48265 hypothetical protein
35	106	2.9 688 2 A33540 transcription factor
36	104	2.9 786 2 A47547 serine proteinase st
37	104	2.9 1095 2 T00349 hypothetical protein
38	106	2.9 1791 2 T02345 hypothetical protein
39	101	2.8 263 4 I67792 probable glucosylcra
40	101	2.8 330 2 PS0079 naringenin-chalcone s
41	101	2.8 424 2 S43560 coiled coil protein B
42	101	2.8 713 2 S58251 probable membrane pro
43	102	2.8 923 2 S44664 ZK370.3 protein - Cae
44	100	2.8 1589 1 R6BYC5 cell division control
45	102	2.8 1645 2 A37792 spectrin beta-H chain

RESULT 1
ENTRY 1
TITLE A49429 #type complete
ORGANISM interleukin-1 beta-converting enzyme homolog CED-3 -
DATE 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994

ACCESSIONS
REFERENCE A49429
#authors Yuan, J.; Shaham, S.; Ledoux, S.; Ellis, H.M.; Horvitz, H.R.
#journal Cell (1993) 73:641-652
#title The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme.
#cross-references MIMD:94061982
#accession A49429
#status Preliminary
#molecule_type DNA
##residues 1-503 #label YUA
##note sequence extracted from NCBI backbone (NCBIN:139825, NCBI:P:139826)

SUMMARY #length 503 #molecular-weight 56616 #checksum 8929

Query Match 99.7%: Score 3591; DB 2; Length 503;
Best Local Similarity 99.8%: Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db	1	MMRDRSLRLERNIMFSSHLLKVDLEVLIAKOVNSONGMINSCTVRRKREIYVA	60
Qy	1	MMRDRSLRLERNIMFSSHLLKVDLEVLIAQVNSONGMINSCTVRRKREIYVA	60
Db	61	VORRGDAFDPAFDALRSTGHEGLAEVLEPLARVSQNVVEECPPSPSHRRSLSPA	120
Qy	61	VORRGDAFDPAFDALRSTGHEGLAEVLEPLARVSQNVVEECPPSPSHRRSLSPA	120
Db	121	GYTSPTVRHDSVSVSFTSYODIYSRARSRSRALHSDRHNSPVPVNAFSPQPS	180
Qy	121	GYTSPTVRHDSVSVSFTSYODIYSRARSRSRALHSDRHNSPVPVNAFSPQPS	180
Db	181	ANSFTGCSLSGYSSSRNSFSKASGPTQYIFHEEDMNFVDAPTISRYVDEKTRNFSS	240
Qy	181	ANSFTGCSLSGYSSSRNSFSKASGPTQYIFHEEDMNFVDAPTISRYVDEKTRNFSS	240
Db	241	PRMCLINNEHFEQMPTRNGTKADKDNLTNFRCGYGVICKDNLTGMLTTRDFAK	300
Qy	241	PRMCLINNEHFEQMPTRNGTKADKDNLTNFRCGYGVICKDNLTGMLTTRDFAK	300
Db	301	HESHGSAIIVLISHGEENVIGVDDIPISTHEIYDLNANANPRLANKPIVFOACRG	360

PI Horvitz HR, Shahan S, Yuan J;
DR WPI: 94-007551/01.
PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death
PT or proliferation
PS Claim 99; : 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -beta convertase (ICE), which converts pro-interleukin-beta to the
CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
CC proteins have an overall amino acid similarity of 28%. The Ced-3
CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA:

Query Match

Best Local Similarity 99.3%; Score 3579; DB 9; Length 503;

Matches 500; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 1 mmrgdrirllernimfssllkvdellvliakgvlnsdngdmnscgtvrekxreivka 60
QY 1 MMQDRRLSLERNIMFSSLLKVDLEVLIAKQVLSNDGDMNSCGTVREKREIYKA 60
DB 61 vgrgqvaafafaydalrctheglaevleplarsvsnavefecpmshrrsralspa 120
QY 61 VGRGQVAFDAFADALRSTCHEGLAEVLEPLARSVDNNAVEFECPMSPASHRRSRALSPA 120
DB 121 gytsptrvhndsvsvsftsygdlysrarsrsralhsdrhysppvnafpsqps 180
QY 121 GYTSPTRVHRDVSYSVSTSYQDIYSRARSRSRALHSDRHNSPPVNAFPSQPS 180
DB 181 anssftgcsllgyssrnsrnsfsasgptylfheedmfvdaptlsrvfdektmyrnfs 240
QY 181 ANSFTGCSLLGYSSRNSRNSFSASGPTYLFHEEDMNFVDAPTISRVEDEKTYRNFS 240
DB 241 prgmcllnehfegmptngtckadkdnltlfrcmgyvickdnltrgmlltrdfak 300
QY 241 PRGMCLINNEHFEQPTNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTRDFAK 300
DB 301 heshdgsalvlshgeenvilgvdipistheiydlinaanaprlanpkivfqacrg 360
QY 301 HESHDDSAVLVLSHGEENVIIIGVDIPISTHEIYDLINAANAPRLANKPIVFQACRG 360
DB 361 errdngfpyldsvdgvpaflirgwdnrdgplfnflgcvrpgvqgwrkkpsgadllrya 420
QY 361 ERRDNGFPYLDSDVDGPVAFILRGWDMNRDGPLNFLGCVRPQVQWRRKPSQADLLIRYA 420
DB 421 ttaqyvswnsargxwflqavcevfsthakmdvvelitevnnkvasafqsgsnllkg 480
QY 421 TTAQYVSWNSARGSMFLQAVCEVFSTHAKMDVVELITEVNNKVAACGQFOTSGSNILLKQ 480
DB 481 mpemtsrllkkfyfpearsav 503
QY 481 MPEMTSRLKKFYFWEARNSAV 503

QY 1 MMRDRSLERIMFSSHLKVDLEVLIAKQVLSNDGDMNSCGTVREKREIYKA 60
 Db 61 vgrgvaafdafydalsrctgheglavleplarsvdsnavefecpmaspashrralspa 120
 QY 61 VQRGVDVAFDAFYDALRSTGHEGLAEVLEPLARSDVSNAVEFECPMSPASHRRALSPLA 120
 Db 121 gytsptrhrdsrvssvstfsygdlysrarsrsralhsdrhnyssppvnafpsgss 180
 QY 121 GYTSPTVRHDSRVSSVSTFSYQDLYSRARSRSRALHSDRHNYSPPVNAFPSGSS 180
 Db 181 ansftgcslygsssrnsfskasgptqylfheedmfvdaplsrvfdektmyrnfs 240
 QY 181 ANSFTGCSLYGSSSRNSFSKASGPTQYLFHEEDMNFVDAPLSRVFDEKTYRNFS 240
 Db 241 prgmcllInnehfemprtrngtkadkdnlnlfrcmgytvckdnltqgmlltrdfak 300
 QY 241 PRGMCLLINNEHFEMPTRNGTKADKDNLTNLFRCMGYTVCKDNLTGRGMLTTRDFAK 300
 Db 301 heshgdsallvllshgeenvllgvdipistheiydlinaanaprlanpkrlvfvqacrg 360
 QY 301 HESHGDSALLVLLSHGEENVLLIGVDIPISTHEIYDLINAAANPRLANPKRLVFOACRG 360
 Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplnflgcvrpgvqvwrrkxpsgadlllrya 420
 QY 361 ERRDNGFPVLDSVDGVPALRRGWDNRDGPLNFLGCVRPQVOOWRRKXPSQADLLIRYA 420
 Db 421 ttagyvsnrnsaragswflgavcevfsthakmdvvellytevnkvasafqtsqsnllkq 480
 QY 421 TTAQVSNRNSARAGSWFIOAVCEVFSTHAKMDVVELLYEVNKKVACGFTSQSGSNLLKQ 480
 Db 481 mpemtsrllkkfyfwpearnasav 503
 QY 481 MPEMTRSLLKKFYFWEARNASAV 503

RESULT 14
 ID R45314 standard; Protein; 503 AA.
 AC R45314;
 DT 01-JUL-1994 (first entry)
 DE Ced-3 mutant A337.
 KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
 KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH Key location/Qualifiers
 FT misc_difference 357
 FT /label= A337
 PN WO9325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; 005705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI, 94-007551/01.
 PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 99; 132pp; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for

CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 SO Sequence 503 AA:
 Query Match 99.3%; Score 3579; DB 9; Length 503;
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;
 Matches 500; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 mmrdrslernmtfsshlkvdlevliaakqvlinsdngdmnscgtvrekreiyka 60
 QY 1 MMRDRSLERIMFSSHLKVDLEVLIAKQVLSNDGDMNSCGTVREKREIYKA 60
 Db 61 vgrgvaafdafydalsrctgheglavleplarsvdsnavefecpmaspashrralspa 120
 QY 61 VQRGVDVAFDAFYDALRSTGHEGLAEVLEPLARSDVSNAVEFECPMSPASHRRALSPLA 120
 Db 121 gytsptrhrdsrvssvstfsygdlysrarsrsralhsdrhnyssppvnafpsgss 180
 QY 121 GYTSPTVRHDSRVSSVSTFSYQDLYSRARSRSRALHSDRHNYSPPVNAFPSGSS 180
 Db 181 ansftgcslygsssrnsfskasgptqylfheedmfvdaplsrvfdektmyrnfs 240
 QY 181 ANSFTGCSLYGSSSRNSFSKASGPTQYLFHEEDMNFVDAPLSRVFDEKTYRNFS 240
 Db 241 prgmcllInnehfemprtrngtkadkdnlnlfrcmgytvckdnltqgmlltrdfak 300
 QY 241 PRGMCLLINNEHFEMPTRNGTKADKDNLTNLFRCMGYTVCKDNLTGRGMLTTRDFAK 300
 Db 301 heshgdsallvllshgeenvllgvdipistheiydlinaanaprlanpkrlvfvqacrg 360
 QY 301 HESHGDSALLVLLSHGEENVLLIGVDIPISTHEIYDLINAAANPRLANPKRLVFOACRG 360
 Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplnflgcvrpgvqvwrrkxpsgadlllrya 420
 QY 361 ERRDNGFPVLDSVDGVPALRRGWDNRDGPLNFLGCVRPQVOOWRRKXPSQADLLIRYA 420
 Db 421 ttagyvsnrnsaragswflgavcevfsthakmdvvellytevnkvasafqtsqsnllkq 480
 QY 421 TTAQVSNRNSARAGSWFIOAVCEVFSTHAKMDVVELLYEVNKKVACGFTSQSGSNLLKQ 480
 Db 481 mpemtsrllkkfyfwpearnasav 503
 QY 481 MPEMTRSLLKKFYFWEARNASAV 503

RESULT 15
 ID R45322 standard; Protein; 503 AA.
 AC R45322;
 DT 01-JUL-1994 (first entry)
 DE Ced-3 mutant S435.
 KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
 KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH Key location/Qualifiers
 FT misc_difference 435
 FT /label= S435
 PN WO9325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; 005705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.

AC R45262; (first entry)
 DT 01-JUL-1994
 DE Ccd-3.
 KW C.elegans; cell death; gene: ccd-3; inhibition; human; parasite;
 KW Interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH Key
 FT peptide location/Qualifiers
 FT 1..372 /note= "Claim 3, inhibitor fragment"
 FT 1..149 /note= "Claim 3, inhibitor fragment"
 FT W09325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DI WPI: 94-007551/01.
 DR N-PSDB: 054401.
 PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 2; Fig 3; 132pp; English.
 CC This sequence is encoded by the C.elegans cell death gene, ccd-3.
 CC Fragments of the amino terminal of this protein act as inhibitors of
 CC ccd-3. The ccd-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ccd-3 may also act as inhibitors of ICE. Human ICE and nematode Ccd-3
 CC proteins have an overall amino acid similarity of 28%. The ccd-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ccd-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC death such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 SO Sequence 503 AA:
 Query Match 99.4%; Score 3583; DB 9; Length 503;
 Best Local Similarity 99.6%; Pred. No. 0.00e+00;
 Matches 501; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 301 heshgdaallvishgeenvigvddipstheiydlinaanaprtaankpkivfgacrg 360
 QY 301 HESHGDAALLVISHGEEVNIIGVDDIPSTHEIYDLINAAANPRANKPKIVFVACAG 360
 DB 361 errdngfpvldsvdyvpaflrrgdwdrpdlfnfigcvrpqvgvvrkpkpsqadallrya 420
 QY 361 ERRDNGFPVLDSVDGVPAFLRRGDMNRDGPFLFNFLCVCBPVOQVWRKPKSOADILIRYA 420
 DB 421 tlaqywsrnarsgsvfigavcevfsthakmdvavllvevkkvasafqtsgsnllqg 480
 QY 421 TLAQYWSRNSAGSWMFIOAVCEVFTSHAKMDVAVLLVEVKKVACGFTSGSNLLQ 480
 DB 481 mpemtsrlkkfyfwpearnasv 503
 QY 481 MPEMTSRLKKFYFWPEARNASV 503
 RESULT 13
 ID R45298 standard; Protein: 503 AA.
 AC R45298;
 DT 01-JUL-1994 (first entry)
 DE Ccd-3 mutant S307.
 KW C.elegans; cell death; gene: ccd-3; inhibition; human; parasite;
 KW Interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH Key
 FT misc location/Qualifiers
 FT 307 /label= S307
 FT W09325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DI WPI: 94-007551/01.
 DR Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 99: 132pp; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ccd-3. These mutations
 CC occur at conserved amino acid residues of the Ccd-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ccd-3. The ccd-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ccd-3 may also act as inhibitors of ICE. Human ICE and nematode Ccd-3
 CC proteins have an overall amino acid similarity of 28%. The ccd-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ccd-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC death such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 SO Sequence 503 AA:
 Query Match 99.3%; Score 3579; DB 9; Length 503;
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;
 Matches 500; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

FT misc_difference 283 /note= "possible mutation site"
FT misc_difference 285 /note= "possible mutation site"
FT misc_difference 286 /note= "possible mutation site"
FT misc_difference 287 /note= "possible mutation site"
FT misc_difference 291 /note= "possible mutation site"
FT misc_difference 298 /note= "possible mutation site"
FT misc_difference 304 /note= "possible mutation site"
FT misc_difference 306 /note= "possible mutation site"
FT misc_difference 307 /note= "possible mutation site"
FT misc_difference 310 /note= "possible mutation site"
FT misc_difference 311 /note= "possible mutation site"
FT misc_difference 314 /note= "possible mutation site"
FT misc_difference 315 /note= "possible mutation site"
FT misc_difference 316 /note= "possible mutation site"
FT misc_difference 321 /note= "possible mutation site"
FT misc_difference 323 /note= "possible mutation site"
FT misc_difference 334 /note= "possible mutation site"
FT misc_difference 339 /note= "possible mutation site"
FT misc_difference 344 /note= "possible mutation site"
FT misc_difference 346 /note= "possible mutation site"
FT misc_difference 349 /note= "possible mutation site"
FT misc_difference 350 /note= "possible mutation site"
FT misc_difference 351 /note= "possible mutation site"
FT misc_difference 356 /note= "possible mutation site"
FT misc_difference 357 /note= "possible mutation site"
FT misc_difference 358 /note= "possible mutation site"
FT misc_difference 359 /note= "C385A active site mutation"
FT misc_difference 360 /note= "possible mutation site"
FT misc_difference 371 /note= "possible mutation site"
FT misc_difference 414 /note= "possible mutation site"
FT misc_difference 414 /note= "possible mutation site"
FT misc_difference 429 /note= "possible mutation site"
FT misc_difference 434 /note= "possible mutation site"
FT misc_difference 435 /note= "possible mutation site"
FT misc_difference 438 /note= "possible mutation site"
FT misc_difference 449 /note= "possible mutation site"
FT misc_difference 452 /note= "possible mutation site"
FT misc_difference 488 /note= "possible mutation site"

```

```

FT /note= "possible mutation site"
FT misc_difference 493 /note= "possible mutation site"
FT misc_difference 496 /note= "possible mutation site"
PN WO9625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
PR 24-FEB-1995; US-394189.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J.
DR WPI: 96-425082/42.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
PS Claim 60: Fig 3: 139pp; English.
CC Inactivation of the Caenorhabditis elegans Ced-3 cell death protein
CC (R98754), pref. through a subseq. of alanine for cysteine at the
CC active site of the protein, provides a mutant Ced-3 protein
CC (R98758) that can prevent programmed cell death in C. elegans
CC when overexpressed using a heat shock promoter. Other mutations
CC at conserved positions can be used to determine amino acid residues
CC important to activity. Corresponding residues in structurally
CC related proteins such as human interleukin-1 beta converting enzyme
CC (R98754) and murine MDD-2 protein (R98756) may then be altered
CC to produce similar effects.
CC (NB. the active site mutation is stated to be C385A rather than
CC C358A in Claim 62).
SQ Sequence 503 AA:

```

Query Match 99.5%; Score 3586; DB 19; Length 503;

Best Local Similarity 99.8%; Pred. No. 0.00e+00; Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Db 1 mmrgdrrslernlmfshlkvdellvllakqvinsdngdmiscgtvrekrrvka 60
QY 1 mmrgdrrslernlmfshlkvdellvllakqvinsdngdmiscgtvrekrrvka 60
Db 61 vqrrgdvafdaftdalstcstheglaevleplarsvsnavefcgmpspashrrslspa 120
QY 61 vqrrgdvafdaftdalstcstheglaevleplarsvsnavefcgmpspashrrslspa 120
Db 61 vqrrgdvafdaftdalstcstheglaevleplarsvsnavefcgmpspashrrslspa 120
QY 61 vqrrgdvafdaftdalstcstheglaevleplarsvsnavefcgmpspashrrslspa 120
Db 121 gysptvhrdsvsvsfsygdlyrsarsrsralhsdrihnysspvaifspqps 180
QY 121 gysptvhrdsvsvsfsygdlyrsarsrsralhsdrihnysspvaifspqps 180
Db 121 gysptvhrdsvsvsfsygdlyrsarsrsralhsdrihnysspvaifspqps 180
QY 121 gysptvhrdsvsvsfsygdlyrsarsrsralhsdrihnysspvaifspqps 180
Db 181 ansftgcsslgyssrnrsfksagptqyifneedmfvadpctsrvtdektmyrnfs 240
QY 181 ansftgcsslgyssrnrsfksagptqyifneedmfvadpctsrvtdektmyrnfs 240
Db 181 ansftgcsslgyssrnrsfksagptqyifneedmfvadpctsrvtdektmyrnfs 240
QY 181 ansftgcsslgyssrnrsfksagptqyifneedmfvadpctsrvtdektmyrnfs 240
Db 241 prgmcllnehefemtrngtkadknltnlfrcmgytvickdnlgtgmlltirdfak 300
QY 241 prgmcllnehefemtrngtkadknltnlfrcmgytvickdnlgtgmlltirdfak 300
Db 241 prgmcllnehefemtrngtkadknltnlfrcmgytvickdnlgtgmlltirdfak 300
QY 241 prgmcllnehefemtrngtkadknltnlfrcmgytvickdnlgtgmlltirdfak 300
Db 301 hesngdsallvllsheenvliqyddipistheiydllaanaaprlanpkxifvgaag 360
QY 301 hesngdsallvllsheenvliqyddipistheiydllaanaaprlanpkxifvgaag 360
Db 301 hesngdsallvllsheenvliqyddipistheiydllaanaaprlanpkxifvgaag 360
QY 301 hesngdsallvllsheenvliqyddipistheiydllaanaaprlanpkxifvgaag 360
Db 361 erdngfvlsvdgvpaflirgwdndgplfnlfgcvrpgvqvrvkksqgdllirya 420
QY 361 erdngfvlsvdgvpaflirgwdndgplfnlfgcvrpgvqvrvkksqgdllirya 420
Db 361 erdngfvlsvdgvpaflirgwdndgplfnlfgcvrpgvqvrvkksqgdllirya 420
QY 361 erdngfvlsvdgvpaflirgwdndgplfnlfgcvrpgvqvrvkksqgdllirya 420
Db 421 ttaqvswrnarsgswfagavcevfshakmdvvelletvknkvaagcftsgsnllkg 480
QY 421 ttaqvswrnarsgswfagavcevfshakmdvvelletvknkvaagcftsgsnllkg 480
Db 481 mpemtsrllkkfyfwearnsav 503
QY 481 mpemtsrllkkfyfwearnsav 503
Db 481 mpemtsrllkkfyfwearnsav 503
QY 481 mpemtsrllkkfyfwearnsav 503

```

RESULT 12
ID R45262 standard: Protein; 503 AA.

QY 361 ERRDNGFPVLDSDVGPAPFLRGMWDRDGPFLNFZGCVRPQVQVWRKKPSQADILIRYA 420
 Db 421 ttaqyvswnsarsgswfifagvcefsthakmdvvevlltevnkkvaccgftsgsnllkq 480
 QY 421 ttaqyvswnsarsgswfifagvcefsthakmdvvevlltevnkkvaccgftsgsnllkq 480
 Db 481 mpemtsrllkkfyfwearnsav 503
 QY 481 MPMTSRLKKFYFWEARNNAV 503

RESULT 10
 ID R53280 standard; Protein; 503 AA.
 AC R53280;
 DT 23-JUN-1994 (first entry)
 DE Ced-3 (G65R).
 KW C. elegans; ced-3; mutant; transcriptional regulation;
 embryogenesis; cell death; hydrophilic; transmembrane; region;
 hydrophobic; mutation; amino acid; substitution; RNA splicing;
 protein synthesis; null phenotype; calcium-binding domain.
 OS Caenorhabditis elegans.
 FH Key
 FT Location/Qualifiers
 FT misc_difference 65
 FT /label= G65R
 FT /note= "from allele n718"
 PN W09325685-A.
 PD 23-DEC-1993.
 PE 14-JUN-1993: U05701.
 PR 12-JUN-1992: US-897788.
 PR 20-NOV-1992: US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI: 94-007542/01.
 DR N-PSDB: 064736.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 develop agents to increase or prevent cell death in organisms
 PS Claim 13; Fig 4; 127pp; English.
 PS The sequences given in R53279-88 are encoded by mutant versions of the
 C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 transcript and was most abundant in embryos, but was also detected in
 larvae and young adults, suggesting that ced-3 is expressed not only
 in cells undergoing cell death. The four largest introns as well as
 sequences 5' of the start codon contain repetitive elements, some of
 which have been characterized in non-coding regions of other C. elegans
 genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 acids in length. Ced-3 is highly hydrophilic with no significant
 hydrophobic region that might be a transmembrane region. One region
 of Ced-3 is very rich in serine. It is thought that this region is
 involved in protein-protein interactions, similar to acid blobs in
 transcription factors. Of the mutations, which occur within the ced-3
 gene, eight of the mutations are missense mutations, two are nonsense
 mutations and two are putative splicing mutations. These mutations
 establish the null phenotype of the ced-3 gene, confirming that ced-3,
 like ced-4, function is not essential for viability. The ced-3 and
 ced-4 gene products may be used to develop agents for treating
 conditions characterized by cell deaths, such as myocardial infarction,
 stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 SQ Sequence 503 AA:

Query Match 99.7%; Score 3591; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmugdrslrllernmtsshlkvdellvllakvlnsdgdmnscgtvrekrrivxa 60
 QY 1 mmqodrsllrllernmtsshlkvdellvllakvlnsdgdmnscgtvrekrrivxa 60
 Db 61 vqrrrdvafafydaistghglaevleplarsvsnavefcmpspshrrsralspa 120
 QY 61 vqrrrdvafafydaistghglaevleplarsvsnavefcmpspshrrsralspa 120

Db 121 gytsptrhdsrvssvstfsygdlysrarsrsralhsdtrhnysspvnafosqps 180
 QY 121 gytsptrhdsrvssvstfsygdlysrarsrsralhsdtrhnysspvnafosqps 180
 Db 181 anssftgcsslgyssrnrsfskaagptqyifhnedmfvdpatisrvfdektmyrnfs 240
 QY 181 ANSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDPATISRVEDEKTYRNFS 240
 Db 241 prgmculinnehpqmrngtrkackdnltlfrcmgyvickdnltlrgmlltrdfak 300
 QY 241 PRGMCLINNEHPQMRNGTRKADKNLTNLFRCMGYVICKDNLTGMLLTRDFAK 300
 Db 301 heshgsaallvlshbeenvllgvdipistheiydlinaanaprllanpkivfagcrg 360
 QY 301 HESHGSALLVLSHBEENVLLGVDDIPISTHEIYDLNAAANAPRLANKPIVFOACRG 360
 Db 361 errdngfpvldsvdygpaflrrgwdnrdgplnflgcvrpqyqvrkkpsqadllirya 420
 QY 361 ERRDNGFPVLDSDVGPAPFLRGMWDRDGPFLNFZGCVRPQVQVWRKKPSQADILIRYA 420
 Db 421 ttaqyvswnsarsgswfifagvcefsthakmdvvevlltevnkkvaccgftsgsnllkq 480
 QY 421 ttaqyvswnsarsgswfifagvcefsthakmdvvevlltevnkkvaccgftsgsnllkq 480
 Db 481 mpemtsrllkkfyfwearnsav 503
 QY 481 MPMTSRLKKFYFWEARNNAV 503

RESULT 11
 ID R98758 standard; Protein; 503 AA.
 AC R98758;
 DT 17-DEC-1996 (first entry)
 DE Nematode ced-3 cell death protein C358A mutant.
 KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
 cell death; apoptosis; neural degeneration; inflammation;
 antiinflammatory.
 OS Synthetic.
 FH Key
 FT Location/Qualifiers
 FT region 78..203
 FT /label= Serine-rich_region
 FT cleavage_site 131
 FT /note= "autocleavage site"
 FT cleavage_site 371
 FT /note= "autocleavage site"
 FT misc_difference 183
 FT /note= "possible mutation site"
 FT misc_difference 234
 FT /note= "possible mutation site"
 FT modified_site 242
 FT /note= "possible mutation site"
 FT misc_difference 246
 FT /note= "possible mutation site"
 FT misc_difference 247
 FT /note= "possible mutation site"
 FT misc_difference 248
 FT /note= "possible mutation site"
 FT misc_difference 250
 FT /note= "possible mutation site"
 FT modified_site 253
 FT /note= "possible mutation site"
 FT misc_difference 259
 FT /note= "possible mutation site"
 FT modified_site 261
 FT /note= "possible mutation site"
 FT misc_difference 265
 FT /note= "possible mutation site"
 FT misc_difference 277
 FT /note= "possible mutation site"
 FT misc_difference 278
 FT /note= "possible mutation site"
 FT misc_difference 280
 FT /note= "possible mutation site"

DT 23-JUN-1994 (first entry)
 DE Ccd-3 (S486f).
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KM embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FI misc_difference 486
 TT /label= "S486f"
 TN /note= "From allele n1163"
 PN MO9325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 PI WPI: 94-007542/01.
 DR N-PSDB: Q64743.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PS develop agents to increase or prevent cell death in organisms
 PS Claim 13: Fig 4: 127pp: English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 CC transcript and was most abundant in embryos, but was also detected in
 CC larvae and young adults, suggesting that ced-3 is expressed not only
 CC in cells undergoing cell death. The four largest introns as well as
 CC sequences 5' of the start codon contain repetitive elements, some of
 CC which have been characterised in non-coding regions of other C. elegans
 CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 CC acids in length. Ced-3 is highly hydrophilic with no significant
 CC hydrophobic region that might be a transmembrane region. One region
 CC of Ced-3 is very rich in serine. It is thought that this region is
 CC involved in protein-protein interactions, similar to acid blobs in
 CC transcription factors. Of the mutations which occur within the ced-3
 CC gene, eight of the mutations are missense mutations, two are nonsense
 CC mutations and two are putative splicing mutations. These mutations
 CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
 CC like ced-4, function is not essential for viability. The ced-3 and
 CC ced-4 gene products may be used to develop agents for treating
 CC conditions characterised by cell deaths, such as myocardial infarction,
 CC stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 CC Sequence 503 AA:
 SQ
 Query Match 99.8%; Score 3595; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 1 mmrgdrrslernimfshhkvdeillevlakvlnsdngdmnscgtvrekreivka 60
 QY 1 mmrgdrrslernimfshhkvdeillevlakvlnsdngdmnscgtvrekreivka 60
 Db 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 QY 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 Db 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 QY 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 Db 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 QY 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 Db 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 QY 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 Db 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 QY 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 Db 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 QY 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 Db 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 QY 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 Db 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 QY 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 Db 301 heshgdsallvllshgeenvilgvdipistheiydlinaanaprlankpkivfagcrg 360
 QY 301 heshgdsallvllshgeenvilgvdipistheiydlinaanaprlankpkivfagcrg 360
 Db 301 HESHGDSALLVLLSHGEENVILGVDIPISTHEIYDLINAANAPRLANKPKIVFQACRG 360
 QY 301 HESHGDSALLVLLSHGEENVILGVDIPISTHEIYDLINAANAPRLANKPKIVFQACRG 360

Db 361 errdngfpvldsvdypafllrrgwdndgplfnflgcvrpyqgvwrkkpsqadillrya 420
 QY 361 ERRDNGFPVLDSDVPALFRRGMDNRDGPJFNFLGCVRPQVQWRRKPKQADILLRYA 420
 Db 421 ttgavsvrnsaragswfifgacvsvfshakmdvveillevknkvaagfsgsgsnllkg 480
 QY 421 ttgavsvrnsaragswfifgacvsvfshakmdvveillevknkvaagfsgsgsnllkg 480
 Db 481 mpemtfllkkfiwpearnsav 503
 QY 481 mpemtfllkkfiwpearnsav 503
 Db 481 mpemtfllkkfiwpearnsav 503
 QY 481 mpemtfllkkfiwpearnsav 503
 RESULT 9
 ID R66770 standard; Protein; 503 AA.
 AC R66770;
 DT 13-SEP-1995 (first entry)
 DE Interleukin-1 beta converting enzyme homolog ced3.
 KW Interleukin-1 beta converting enzyme; homolog ced3;
 OS Caenorhabditis elegans.
 PN WO9500160-A.
 PD 05-JAN-1995.
 PF 10-JUN-1994; U06630.
 PR 24-JUN-1993; US-080850.
 PA (GEHO) GEN HOSPITAL CORP.
 PI Miura M, Yuan J;
 PI WPI: 95-051742/07.
 DR N-PSDB: Q79970.
 PT Promoting or preventing programmed cell death in vertebrate cells
 PT - by inhibiting the activity of interleukin-1 beta converting
 PT enzyme.
 PS Example 1; Fig 2C: 116pp: English.
 CC Q79970 encodes R66770 interleukin-1 beta converting enzyme homolog
 CC ced3, increasing ced3s enzymatic activity can promote the
 CC programmed cell death of cancer cells (pref. those overexpressing
 CC the bcl-2 oncogene) this can be used as the basis of a new cancer
 CC treatment. Alternatively by reducing ced3s enzymatic activity
 CC programmed cell death can be inhibited, this may be useful in the
 CC development of new cell lines which remain viable in culture for
 CC extended or indefinite periods, independent of growth factors.
 CC Sequence 503 AA:
 SQ
 Query Match 99.7%; Score 3593; DB 13; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 1 mmrgdrrslernimfshhkvdeillevlakvlnsdngdmnscgtvrekreivka 60
 QY 1 mmrgdrrslernimfshhkvdeillevlakvlnsdngdmnscgtvrekreivka 60
 Db 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 QY 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 Db 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 QY 61 vqrgvafafafdalrstrctgheglavleplarsvdsnavefecpmpashrrsralpa 120
 Db 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 QY 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 Db 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 QY 121 gytsptrhrdsvsvsftsdydylstrarsrsralhsdrrhysppnaafspqss 180
 Db 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 QY 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 Db 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 QY 181 anssftgcslgysssrnsfskaspqcyifhfeedmfvadpatsrvfdektmyrnfs 240
 Db 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 QY 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 Db 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 QY 241 prgmcllilnehfegmptrngtkadkdnltlfrmgtyvickdnltrgmlltrdfak 300
 Db 301 heshgdsallvllshgeenvilgvdipistheiydlinaanaprlankpkivfagcrg 360
 QY 301 HESHGDSALLVLLSHGEENVILGVDIPISTHEIYDLINAANAPRLANKPKIVFQACRG 360
 Db 361 errdngfpvldsvdypafllrrgwdndgplfnflgcvrpyqgvwrkkpsqadillrya 420

CC involved in protein-protein interactions, similar to acid blobs in
 CC transcription factors. Of the mutations which occur within the ced-3
 CC gene, eight of the mutations are missense mutations, two are nonsense
 CC mutations and two are putative splicing mutations. These mutations
 CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
 CC like ced-4, function is not essential for viability. The ced-3 and
 CC ced-4 gene products may be used to develop agents for treating
 CC conditions characterised by cell deaths, such as myocardial infarction,
 CC stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 CC Sequence 503 AA:

Query Match 99.8%; Score 3596; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db 1 mmrgdrrllernimmfshlkvdeifevliakvlnsdngdmnscgtvrekreivka 60
    |||||||
Qy 1 MMRODRSLERINIMFSSHLKVDLEVLIAKQVLSNNGDMINSCGTVERREIYKA 60
    |||||||
Db 61 vrrigvafdaifdalrstgheglaevlleplarsvdsnavfecpmshrralspa 120
    |||||||
Qy 61 VORRGVAADFADALRSTGHEGLAEVLLEPLARSDVSNAVEFCPMSPASHRRSRLSPA 120
    |||||||
Db 121 gysptvrhdsavsvsftsdygdlysrarsrralhsdtrhysppvnafpsqps 180
    |||||||
Qy 121 GYSPTRVHRDSSVSVSFTSYODIYSRARSRRSRLHSSDRHNSSPPVNAFPQPS 180
    |||||||
Db 181 ansftgcslgyssrnsrfskaspqlyifheedmfvdapltisrvidekmyrnfs 240
    |||||||
Qy 181 ANSFTGCSSLGYSSRNSRFSKASGPTYIFHEEDMNFVDAPLTISRVIDEKTMYRNFS 240
    |||||||
Db 241 prgmcllnehefemprngtkadkdnlnlfrcmgyvickdnltgymlltirdfak 300
    |||||||
Qy 241 PRGMCLINNEHEFQMPTRNGTKADKDNLTNLFRCMGYVICKDNLTGGMLLTIRDFAK 300
    |||||||
Db 301 heshgdsallvlshgeenvilgvdipistheiyqllnaanprlanpkivfvacag 360
    |||||||
Qy 301 HESHGDSALLVLSHGEENVIIIGVDIPISTHEIYDLNNAANPRLANPKIVFVQACAG 360
    |||||||
Db 361 errdngfpvldsvdygpaflrrgwdnrdgplfnflgcvrpqvgvwrkkpsgadllira 420
    |||||||
Qy 361 ERRDNGFPVLDSVDYGPALRRGWDNRDGPLFNFLGCVRPQVOQVWRKKPSQADILIRXA 420
    |||||||
Db 421 ttagyvrwnsargswfifagvcevfsthakmdvvellevnkkaacgftsggsnllkq 480
    |||||||
Qy 421 TTAGYVSRNSARGSWFIOAVCEVFSTHAKMDVVELLEVNKKAACGFTSGGSNLLKQ 480
    |||||||
Db 481 mpemtsrllkkfyfpearnsav 503
    |||||||
Qy 481 MEPMTSRLKKFYFPEARNSAV 503
    |||||||

```

RESULT 7
 ID R53287 standard; Protein: 503 AA.
 AC R53287;
 DT 23-JUN-1994 (first entry)
 DE Ced-3 (E483K).
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT misc_difference 483
 FT /Label= E483K
 FT /note= "From allele n2426"
 PN MO9325685-A.
 PD 23-DEC-1993.
 PR 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Horvitz HR, Shahan S, Yuan J;
 DR WPI: 94-007542/01.
 DR N-PSDB: 064744.
 PR Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PR develop agents to increase or prevent cell death in organisms
 PS Claim 13: Fig 4; 127pp; English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 CC transcript and was most abundant in embryos, but was also detected in
 CC larvae and young adults, suggesting that ced-3 is expressed not only
 CC in cells undergoing cell death. The four largest introns as well as
 CC sequences 5' of the start codon contain repetitive elements, some of
 CC which have been characterised in non-coding regions of other C. elegans
 CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 CC acids in length. Ced-3 is highly hydrophilic with no significant
 CC hydrophobic region that might be a transmembrane region. One region
 CC of Ced-3 is very rich in serine. It is thought that this region is
 CC involved in protein-protein interactions, similar to acid blobs in
 CC transcription factors. Of the mutations which occur within the ced-3
 CC gene, eight of the mutations are missense mutations, two are nonsense
 CC mutations and two are putative splicing mutations. These mutations
 CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
 CC like ced-4, function is not essential for viability. The ced-3 and
 CC ced-4 gene products may be used to develop agents for treating
 CC conditions characterised by cell deaths, such as myocardial infarction,
 CC stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 CC Sequence 503 AA:

Query Match 99.8%; Score 3596; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Db 1 mmrgdrrllernimmfshlkvdeifevliakvlnsdngdmnscgtvrekreivka 60
    |||||||
Qy 1 MMRODRSLERINIMFSSHLKVDLEVLIAKQVLSNNGDMINSCGTVERREIYKA 60
    |||||||
Db 61 vrrigvafdaifdalrstgheglaevlleplarsvdsnavfecpmshrralspa 120
    |||||||
Qy 61 VORRGVAADFADALRSTGHEGLAEVLLEPLARSDVSNAVEFCPMSPASHRRSRLSPA 120
    |||||||
Db 121 gysptvrhdsavsvsftsdygdlysrarsrralhsdtrhysppvnafpsqps 180
    |||||||
Qy 121 GYSPTRVHRDSSVSVSFTSYODIYSRARSRRSRLHSSDRHNSSPPVNAFPQPS 180
    |||||||
Db 181 ansftgcslgyssrnsrfskaspqlyifheedmfvdapltisrvidekmyrnfs 240
    |||||||
Qy 181 ANSFTGCSSLGYSSRNSRFSKASGPTYIFHEEDMNFVDAPLTISRVIDEKTMYRNFS 240
    |||||||
Db 241 prgmcllnehefemprngtkadkdnlnlfrcmgyvickdnltgymlltirdfak 300
    |||||||
Qy 241 PRGMCLINNEHEFQMPTRNGTKADKDNLTNLFRCMGYVICKDNLTGGMLLTIRDFAK 300
    |||||||
Db 301 heshgdsallvlshgeenvilgvdipistheiyqllnaanprlanpkivfvacag 360
    |||||||
Qy 301 HESHGDSALLVLSHGEENVIIIGVDIPISTHEIYDLNNAANPRLANPKIVFVQACAG 360
    |||||||
Db 361 errdngfpvldsvdygpaflrrgwdnrdgplfnflgcvrpqvgvwrkkpsgadllira 420
    |||||||
Qy 361 ERRDNGFPVLDSVDYGPALRRGWDNRDGPLFNFLGCVRPQVOQVWRKKPSQADILIRXA 420
    |||||||
Db 421 ttagyvrwnsargswfifagvcevfsthakmdvvellevnkkaacgftsggsnllkq 480
    |||||||
Qy 421 TTAGYVSRNSARGSWFIOAVCEVFSTHAKMDVVELLEVNKKAACGFTSGGSNLLKQ 480
    |||||||
Db 481 mpemtsrllkkfyfpearnsav 503
    |||||||
Qy 481 MEPMTSRLKKFYFPEARNSAV 503
    |||||||

```

RESULT 8
 ID R53288 standard; Protein: 503 AA.
 AC R53288;

QY 241 PRGMCLINNEHEPQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDEAK 300
 Db 301 heshgdsailv1lshgeenvi1gvdip1stheiyd1l1naanaapl1ankpk1vfvagcs 360
 QY 301 HESHGDSAILV1LSHGEENV1IGVDIP1STHEIYDL1NAANAAP1LANKPK1VFOACRG 360
 Db 361 errdngfpy1dsvdgy1paf1rrgwdndqpl1fn1jgcvrpyqgvwv1kksqad1l1rya 420
 QY 361 ERRDNGFPY1DSVDGY1PAFLRRGMDNRDGP1FNLGCVRPQVQVWVKKRPSQAD1L1RYA 420
 Db 421 tlaqyvswnsarsgsw1figavcevf1sthakmdv1evn1kvvacg1fqtsgsn1lkq 480
 QY 421 TTAQYVSWNSARSGSW1FOAVCEVFSTHAKMDV1ELTEVNKKVACG1FQTSOGSN1LKQ 480
 Db 481 mpemtsr1l1kk1fyf1wpear1nsav 503
 QY 481 MPEMTR1LKK1FYFWPEARNSAV 503

RESULT 5

ID R53281 standard; Protein; 503 AA.
 AC R53281;
 DT 23-JUN-1994 (first entry)
 DE Ced-3 (G360S).
 KM C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KM embryogenesis; cell death; hydrophilic; transmembrane; region;
 KM hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KM protein synthesis; null phenotype; calcium-binding domain.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT misc-difference 360
 FT /label= "G360S"
 FT /note= "From allele n2433"

PN MO9325685-A.
 PD 23-DEC-1993. U05701.
 PF 14-JUN-1993: U05701.
 PR 12-JUN-1992: US-897788.
 PR 20-NOV-1992: US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 PI WPI: 94-007542/01.
 DR N-PSDB: 064737.
 DR 27 isolated C elegans cell death genes ced-3 and ced-4 - used to
 develop agents to increase or prevent cell death in organisms
 PS Claim 13: Fig 4; 127P; English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 CC transcript and was most abundant in embryos, but was also detected in
 CC larvae and young adults, suggesting that ced-3 is expressed not only
 CC in cells undergoing cell death. The four largest introns as well as
 CC sequences 5' of the start codon contain repetitive elements, some of
 CC which have been characterised in non-coding regions of other C. elegans
 CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 CC acids in length. Ced-3 is highly hydrophilic with no significant
 CC hydrophobic region that might be a transmembrane region. One region
 CC of Ced-3 is very rich in serine. It is thought that this region is
 CC involved in protein-protein interactions, similar to acid blobs in
 CC transcription factors. Of the mutations, which occur within the ced-3
 CC gene, eight of the mutations are missense mutations, two are nonsense
 CC mutations and two are putative splicing mutations. These mutations
 CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
 CC like ced-4, function is not essential for viability. The ced-3 and
 CC ced-4 gene products may be used to develop agents for treating
 CC conditions characterised by cell deaths, such as myocardial infarction,
 CC stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 SQ Sequence 503 AA;

Query Match 99.8%; Score 3597; DB 9; Length 503;
 Best Local Similarity 99.8%; P-Id. No. 0.00e+00;
 Matches 502; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 1 mmrgrt1l1ern1mmfsh1k1vde1lev1l1ak1y1nsdng1m1nsc1g1v1r1e1r1v1a 60

QY 1 MMRDRSL1ERN1MMFSSHLK1VDE1LEV1L1A1OV1NSD1G1M1N1SC1G1V1R1E1R1V1A 60
 Db 61 vgrgrd1afad1al1rst1ghe1laev1leplarsvsnavefecm1psash1rsal1spa 120
 QY 61 VGRGRDAFADAFYALRSTGHEGLAEVLEPLARSVSNAVEFECMPSPASHRRSRLSPA 120
 Db 121 gyspt1rvh1rdsvs1vss1fcs1yqdl1yars1rs1ral1hsd1th1y1ss1p1vna1f1sp1gs 180
 QY 121 GYSPTRVHRDVSVS1VSS1FCS1YQD1YARS1RS1RAL1HSD1TH1Y1SS1P1VNA1F1SP1GS 180
 Db 181 anss1fcs1g1y1ss1rns1fs1kas1gpl1y1lf1f1hed1n1f1dapt1sv1f1d1ek1m1yr1f1ss 240
 QY 181 ANSSFTGCS1GLY1SS1RNS1FS1KAS1GPL1Y1LF1F1HED1N1F1DAPT1SV1F1D1EK1M1YR1F1SS 240
 Db 241 prgmcl1inne1he1qmp1trng1tkad1n1l1n1frcm1gy1tv1ick1dnl1trg1ml1t1r1d1e1ak 300
 QY 241 PRGMCL1INNEHEPQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDEAK 300
 Db 301 heshgdsailv1lshgeenvi1gvdip1stheiyd1l1naanaapl1ankpk1vfvagcs 360
 QY 301 HESHGDSAILV1LSHGEENV1IGVDIP1STHEIYDL1NAANAAP1LANKPK1VFOACRG 360
 Db 361 errdngfpy1dsvdgy1paf1rrgwdndqpl1fn1jgcvrpyqgvwv1kksqad1l1rya 420
 QY 361 ERRDNGFPY1DSVDGY1PAFLRRGMDNRDGP1FNLGCVRPQVQVWVKKRPSQAD1L1RYA 420
 Db 421 tlaqyvswnsarsgsw1figavcevf1sthakmdv1evn1kvvacg1fqtsgsn1lkq 480
 QY 421 TTAQYVSWNSARSGSW1FOAVCEVFSTHAKMDV1ELTEVNKKVACG1FQTSOGSN1LKQ 480
 Db 481 mpemtsr1l1kk1fyf1wpear1nsav 503
 QY 481 MPEMTR1LKK1FYFWPEARNSAV 503

RESULT 6

ID R53279 standard; Protein; 503 AA.
 AC R53279;
 DT 23-JUN-1994 (first entry)
 DE Ced-3 (L27F).
 KM C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KM embryogenesis; cell death; hydrophilic; transmembrane; region;
 KM hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KM protein synthesis; null phenotype; calcium-binding domain.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT misc-difference 27
 FT /label= "L27F"
 FT /note= "From allele n1040"

PN MO9325685-A.
 PD 23-DEC-1993. U05701.
 PF 14-JUN-1993: U05701.
 PR 12-JUN-1992: US-897788.
 PR 20-NOV-1992: US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 PI WPI: 94-007542/01.
 DR N-PSDB: 064735.
 DR 27 isolated C elegans cell death genes ced-3 and ced-4 - used to
 develop agents to increase or prevent cell death in organisms
 PS Claim 13: Fig 4; 127P; English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 CC transcript and was most abundant in embryos, but was also detected in
 CC larvae and young adults, suggesting that ced-3 is expressed not only
 CC in cells undergoing cell death. The four largest introns as well as
 CC sequences 5' of the start codon contain repetitive elements, some of
 CC which have been characterised in non-coding regions of other C. elegans
 CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 CC acids in length. Ced-3 is highly hydrophilic with no significant
 CC hydrophobic region that might be a transmembrane region. One region
 CC of Ced-3 is very rich in serine. It is thought that this region is

PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI: 94-007542/01.
 DR N-PSDB: 064742.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 develop agents to increase or prevent cell death in organisms
 PS Claim 13: Fig 4: 127pp: English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 transcript and was most abundant in embryos, but was also detected in
 larvae and young adults, suggesting that ced-3 is expressed not only
 in cells undergoing cell death. The four largest introns as well as
 sequences 5' of the start codon contain repetitive elements, some of
 which have been characterised in non-coding regions of other C. elegans
 genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 acids in length. Ced-3 is highly hydrophilic with no significant
 hydrophobic region that might be a transmembrane region. One region
 of Ced-3 is very rich in serine. It is thought that this region is
 involved in protein-protein interactions, similar to acid blobs in
 transcription factors. Of the mutations which occur within the ced-3
 gene, eight of the mutations are missense mutations, two are nonsense
 mutations and two are putative splicing mutations. These mutations
 establish the null phenotype of the ced-3 gene, confirming that ced-3,
 like ced-4, function is not essential for viability. The ced-3 and
 ced-4 gene products may be used to develop agents for treating
 conditions characterised by cell deaths, such as myocardial infarction,
 stroke, degenerative disease, traumatic brain injury, hypoxia,
 pathogenic infection, aging or hair loss.
 CC Sequence 503 AA;

Query Match 99.9%; Score 3599; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 mmrgdrrslernlmfshlkvdellvllakqvlinsdngdmnscgtvrekrrrelvka 60
 QY 1 MMRDRSLRLERNIMFSSHLKVDLEVLIAQVINSNDGDMINSGTVERREKRELIVKA 60
 DB 61 vrrrgdvaafafdalstgheglaeyleplarsvdsnavefecpmaspashrrsralspa 120
 QY 61 VÖRRGDVAFAFADALSTGHEGLAEYLEPLARSVDNNAVEFECPMSPASHRRSRAISPA 120
 DB 121 gysptvrhdsvsfscysqdiysrarsrslalhsddrhysppvnaifpsqps 180
 QY 121 GYSPTRVHRDVSVSFSTYQDIYSRARSRSRALHSSDHNHNSPPVNAFPSPQSS 180
 DB 181 ansffgcslgyssrnrsfksasptqylfneedmfvadpjlsvfidektmyrnfas 240
 QY 181 ANSFFGCSSLGYSSRNRSFKASGPTQYLFNEEDMNFVADPTISRVEDEKMYRNFS 240
 DB 241 prgmcllnehegmptngtkadnlnlfrmgycvickdnltrgmlltridfex 300
 QY 241 PRGMCLLINNEHEGMPTNGTKADNLTNLFRCMGYVICKDNLTRGMMLTTRIDFEAK 300
 DB 301 heshgdsailvlshegnvllqvddipistheiydlinaaanprlanpkivfoacrg 360
 QY 301 HESHGDSAILVLSHGEENVIIQVDIPRSTHEITYDLINAANPRLANPKIVFOACRG 360
 DB 361 errdingfpyldsvdypafllrrgwdndrplfnflgcvrpgvqvwrrkpsqadllirya 420
 QY 361 ERRDNGFPYLDSDYVPAFLRRGMDNRDPLFNFLGCVRPQVOVWRKRPSSADLLIRYA 420
 DB 421 ttaqvswnsaragwffiaavevsthnkmdvwelllevnkxvaacgftsqsnllkq 480
 QY 421 TTAQVSWNSARAGWFFIAAVEVSTHNKMDVWELLLEVNKXVACGFTSQSGNLLKQ 480
 DB 481 mpemtsrlkkfyfwearnsav 503
 QY 481 MPEMTRSRLKKFYFWEARNSAV 503

RESULT 4
 ID R53286 standard; Protein: 503 AA.
 AC R53286;
 DT 23-JUN-1994 (first entry)
 DE Ced-3 (A466v).
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT misc_difference 466
 FT /label= A466v
 FT /note= "From allele n2430"
 PN M09325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI: 94-007542/01.
 DR N-PSDB: 064743.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 develop agents to increase or prevent cell death in organisms
 PS Claim 13: Fig 4: 127pp: English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 transcript and was most abundant in embryos, but was also detected in
 larvae and young adults, suggesting that ced-3 is expressed not only
 in cells undergoing cell death. The four largest introns as well as
 sequences 5' of the start codon contain repetitive elements, some of
 which have been characterised in non-coding regions of other C. elegans
 genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 acids in length. Ced-3 is highly hydrophilic with no significant
 hydrophobic region that might be a transmembrane region. One region
 of Ced-3 is very rich in serine. It is thought that this region is
 involved in protein-protein interactions, similar to acid blobs in
 transcription factors. Of the mutations which occur within the ced-3
 gene, eight of the mutations are missense mutations, two are nonsense
 mutations and two are putative splicing mutations. These mutations
 establish the null phenotype of the ced-3 gene, confirming that ced-3,
 like ced-4, function is not essential for viability. The ced-3 and
 ced-4 gene products may be used to develop agents for treating
 conditions characterised by cell deaths, such as myocardial infarction,
 stroke, degenerative disease, traumatic brain injury, hypoxia,
 pathogenic infection, aging or hair loss.
 CC Sequence 503 AA;

Query Match 99.9%; Score 3599; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 mmrgdrrslernlmfshlkvdellvllakqvlinsdngdmnscgtvrekrrrelvka 60
 QY 1 MMRDRSLRLERNIMFSSHLKVDLEVLIAQVINSNDGDMINSGTVERREKRELIVKA 60
 DB 61 vrrrgdvaafafdalstgheglaeyleplarsvdsnavefecpmaspashrrsralspa 120
 QY 61 VÖRRGDVAFAFADALSTGHEGLAEYLEPLARSVDNNAVEFECPMSPASHRRSRAISPA 120
 DB 121 gysptvrhdsvsfscysqdiysrarsrslalhsddrhysppvnaifpsqps 180
 QY 121 GYSPTRVHRDVSVSFSTYQDIYSRARSRSRALHSSDHNHNSPPVNAFPSPQSS 180
 DB 181 ansffgcslgyssrnrsfksasptqylfneedmfvadpjlsvfidektmyrnfas 240
 QY 181 ANSFFGCSSLGYSSRNRSFKASGPTQYLFNEEDMNFVADPTISRVEDEKMYRNFS 240
 DB 241 prgmcllnehegmptngtkadnlnlfrmgycvickdnltrgmlltridfex 300
 QY 241 PRGMCLLINNEHEGMPTNGTKADNLTNLFRCMGYVICKDNLTRGMMLTTRIDFEAK 300

CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
CC Sequence 503 AA.

Query Match 100.0%; Score 3603; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 mmrqrdrsllemnimmfshlkvdeilevliakvlnsdngdmnscgtvrekreivya 60
CC |MMQDRSLLENNIMFSSHLKVDLEVLAKQVLSNDGDMNISCCTVREKREIYKA 60
QY 1 MMQDRSLLENNIMFSSHLKVDLEVLAKQVLSNDGDMNISCCTVREKREIYKA 60
DB 61 vqrrgvaafafdafrstgheglaeleplarsvdsnavefecpmaspshrslspa 120
CC |VQRGVAFAFADFALRSTGHEGLAELEPLARSDSNAVEFECPMASPHRSRLSPA 120
QY 61 VQRGVAFAFADFALRSTGHEGLAELEPLARSDSNAVEFECPMASPHRSRLSPA 120
DB 121 gytsptrvhrdsvsvsfstygdyrsarsrsralhsdrhnyssppvnaifspqps 180
CC |GYSPTVRHRDSVSVSFTSYODIYSRARSRSRALHSSDRHNYSSPPVNAIFSPQPS 180
QY 121 GYSPTVRHRDSVSVSFTSYODIYSRARSRSRALHSSDRHNYSSPPVNAIFSPQPS 180
DB 181 ansstfgcslygssrnrsfskaspqiyifneedmfvadpilsrvfdektmyrnfs 240
CC |ANSSFTGCSLYGSSRNRSFSKASPTQYIFHEEDMNFVADPILSRVFEKTYRNFS 240
QY 181 ANSSFTGCSLYGSSRNRSFSKASPTQYIFHEEDMNFVADPILSRVFEKTYRNFS 240
DB 241 prgmcllnehefegmptnrgtkadkdnlnlfrcmgytvlckdnltrgmlltridfak 300
CC |PRGMCLLNEHEFEGMPTNRTGKADKDNLNLFRCMGYTVLCKDNLTRGMLLTRIDFAK 300
QY 241 PRGMCLLNEHEFEGMPTNRTGKADKDNLNLFRCMGYTVLCKDNLTRGMLLTRIDFAK 300
DB 301 heshgdsallvllshgeenvllgyvdipistheiydlinaaanprlanpkivfvagcrg 360
CC |HESHGDSALLVLLSHGEENVLLIGVDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
QY 301 HESHGDSALLVLLSHGEENVLLIGVDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
DB 361 errdngfplvdsdgvpafllrrgwdndrgplfnflgcvrpgvqgvwrkksqadllirya 420
CC |ERRDNGFPLVDSVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
QY 361 ERRDNGFPLVDSVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
DB 421 ttaqvswrnarsargswfagavcefsthakmdvveltevnkkvaagcftsgsgnllkq 480
CC |TTAQVSWRNARSARGSWFAGAVCEVFSHAKMDVVELTEVNKKVACGFTSGSGNLLKQ 480
QY 421 TTAQVSWRNARSARGSWFAGAVCEVFSHAKMDVVELTEVNKKVACGFTSGSGNLLKQ 480
DB 481 mpemtsrllkkfyfwpearnsav 503
CC |MPEMTSRLKKFYFWPEARNSAV 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 2
ID R98754 standard; Protein: 503 AA.
AC R98754;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 cell death protein.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory.
OS Caenorhabditis elegans.
FH Key
FH region Location/Qualifiers
FT 78..205
FT /label= Serine-rich_region
FT cleavage_site 131
FT /note= "autocleavage site"
FT cleavage_site 371
FT /note= "autocleavage site"
FT active_site 358
FT /note= "active site Cysteine residue"

W09625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996: U02473.
PR 24-FEB-1995: US-394189.
PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR N-PSDB: T38196.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins

PT useful to treat inflammation and diseases characterised by cell
PT death
PS Claim 18: Fig 3; 139pp; English.
CC The Caenorhabditis elegans Ced-3 protein (R98754) is encoded
CC by the ced-3 cell death gene (T38196). It is structurally similar
CC to human interleukin-1 beta converting enzyme (ICE) (R98755); the
CC 2 proteins have 28% amino acid identity (higher in the C-terminal
CC regions critical for activity), suggesting that Ced-3 may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death protein. Ced-3 proteins can therefore
CC be used in prodn. of novel drugs for enhancing or inhibiting the
CC activity of ICE for the treatment of inflammatory diseases and/or
CC diseases caused by cell death. Novel inhibitors of ced-3 activity
CC include portions of Ced-3. Ced-3 with a C360A subtn. (see also
CC R98758) can prevent programmed cell death in C. elegans.
SQ Sequence 503 AA.

Query Match 100.0%; Score 3603; DB 19; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 mmrqrdrsllemnimmfshlkvdeilevliakvlnsdngdmnscgtvrekreivya 60
CC |MMQDRSLLENNIMFSSHLKVDLEVLAKQVLSNDGDMNISCCTVREKREIYKA 60
QY 1 MMQDRSLLENNIMFSSHLKVDLEVLAKQVLSNDGDMNISCCTVREKREIYKA 60
DB 61 vqrrgvaafafdafrstgheglaeleplarsvdsnavefecpmaspshrslspa 120
CC |VQRGVAFAFADFALRSTGHEGLAELEPLARSDSNAVEFECPMASPHRSRLSPA 120
QY 61 VQRGVAFAFADFALRSTGHEGLAELEPLARSDSNAVEFECPMASPHRSRLSPA 120
DB 121 gytsptrvhrdsvsvsfstygdyrsarsrsralhsdrhnyssppvnaifspqps 180
CC |GYSPTVRHRDSVSVSFTSYODIYSRARSRSRALHSSDRHNYSSPPVNAIFSPQPS 180
QY 121 GYSPTVRHRDSVSVSFTSYODIYSRARSRSRALHSSDRHNYSSPPVNAIFSPQPS 180
DB 181 ansstfgcslygssrnrsfskaspqiyifneedmfvadpilsrvfdektmyrnfs 240
CC |ANSSFTGCSLYGSSRNRSFSKASPTQYIFHEEDMNFVADPILSRVFEKTYRNFS 240
QY 181 ANSSFTGCSLYGSSRNRSFSKASPTQYIFHEEDMNFVADPILSRVFEKTYRNFS 240
DB 241 prgmcllnehefegmptnrgtkadkdnlnlfrcmgytvlckdnltrgmlltridfak 300
CC |PRGMCLLNEHEFEGMPTNRTGKADKDNLNLFRCMGYTVLCKDNLTRGMLLTRIDFAK 300
QY 241 PRGMCLLNEHEFEGMPTNRTGKADKDNLNLFRCMGYTVLCKDNLTRGMLLTRIDFAK 300
DB 301 heshgdsallvllshgeenvllgyvdipistheiydlinaaanprlanpkivfvagcrg 360
CC |HESHGDSALLVLLSHGEENVLLIGVDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
QY 301 HESHGDSALLVLLSHGEENVLLIGVDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
DB 361 errdngfplvdsdgvpafllrrgwdndrgplfnflgcvrpgvqgvwrkksqadllirya 420
CC |ERRDNGFPLVDSVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
QY 361 ERRDNGFPLVDSVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
DB 421 ttaqvswrnarsargswfagavcefsthakmdvveltevnkkvaagcftsgsgnllkq 480
CC |TTAQVSWRNARSARGSWFAGAVCEVFSHAKMDVVELTEVNKKVACGFTSGSGNLLKQ 480
QY 421 TTAQVSWRNARSARGSWFAGAVCEVFSHAKMDVVELTEVNKKVACGFTSGSGNLLKQ 480
DB 481 mpemtsrllkkfyfwpearnsav 503
CC |MPEMTSRLKKFYFWPEARNSAV 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 3
ID R53285 standard; Protein: 503 AA.
AC R53285;
DT 23-JUN-1994 (first entry)
DE Ced-3 (A449V).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key
FH Location/Qualifiers
FT misc_difference 449
FT /label= A449V
FT /note= "from allele n1164"

W09325685-A.

 WORMS
 ***** (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

MPsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:42:11 1999; Maspar time 19.74 Seconds

Tabular output not generated. 541.993 Million cell updates/sec

Title: >US-08-287-669-19

Description: (1-503) from US08287669.pep

Sequence: 1 MMQRDRSRSLERNIMFSSH.....MTRLLKFFWFPARNSAV 503

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseq35
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 35.995; Variance 158.041; scale 0.228

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	3603	100.0	503	9	R47466	Ced-3	0.00e+00
2	3603	100.0	503	19	R98754	Nematode Ced-3 cell d	0.00e+00
3	3599	99.9	503	9	R53285	Ced-3 (A449V).	0.00e+00
4	3599	99.9	503	9	R53286	Ced-3 (A46V).	0.00e+00
5	3597	99.8	503	9	R53281	Ced-3 (G360S).	0.00e+00
6	3596	99.8	503	9	R53279	Ced-3 (L27F).	0.00e+00
7	3596	99.8	503	9	R53287	Ced-3 (E483K).	0.00e+00
8	3595	99.7	503	13	R53288	Ced-3 (S486F).	0.00e+00
9	3593	99.7	503	13	R66770	Interleukin-1 beta co	0.00e+00
10	3591	99.7	503	9	R53280	Ced-3 (G65R).	0.00e+00
11	3586	99.5	503	19	R98758	Nematode Ced-3 cell d	0.00e+00
12	3583	99.4	503	9	R45262	Ced-3	0.00e+00
13	3579	99.3	503	9	R45298	Ced-3 mutant S307.	0.00e+00
14	3579	99.3	503	9	R45314	Ced-3 mutant A357.	0.00e+00
15	3579	99.3	503	9	R45322	Ced-3 mutant S435.	0.00e+00
16	3579	99.3	503	9	R45324	Ced-3 mutant A449.	0.00e+00

17	3579	99.3	503	9	R45301	Ced-3 mutant S314.	0.00e+00
18	3579	99.3	503	9	R45276	Ced-3 mutant S183.	0.00e+00
19	3578	99.3	503	9	R45282	Ced-3 mutant N250.	0.00e+00
20	3578	99.3	503	9	R45291	Ced-3 mutant N285.	0.00e+00
21	3578	99.3	503	9	R45293	Ced-3 mutant T287.	0.00e+00
22	3578	99.3	503	9	R45307	Ced-3 mutant N339.	0.00e+00
23	3577	99.3	503	9	R45318	Ced-3 mutant D371.	0.00e+00
24	3577	99.3	503	9	R45319	Ced-3 mutant D414.	0.00e+00
25	3577	99.3	503	9	R45297	Ced-3 mutant D306.	0.00e+00
26	3577	99.3	503	9	R45286	Ced-3 mutant D265.	0.00e+00
27	3576	99.3	503	9	R45280	Ced-3 mutant I247.	0.00e+00
28	3576	99.3	503	9	R45285	Ced-3 mutant G261.	0.00e+00
29	3576	99.3	503	9	R45321	Ced-3 mutant G434.	0.00e+00
30	3576	99.3	503	9	R45287	Ced-3 mutant G277.	0.00e+00
31	3576	99.3	503	9	R45290	Ced-3 mutant K283.	0.00e+00
32	3576	99.3	503	9	R45306	Ced-3 mutant I334.	0.00e+00
33	3576	99.3	503	9	R45317	Ced-3 mutant K360.	0.00e+00
34	3576	99.3	503	9	R45300	Ced-3 mutant V311.	0.00e+00
35	3576	99.3	503	9	R45313	Ced-3 mutant Q356.	0.00e+00
36	3576	99.3	503	9	R45323	Ced-3 mutant I438.	0.00e+00
37	3576	99.3	503	9	R45303	Ced-3 mutant V280.	0.00e+00
38	3576	99.3	503	9	R45289	Ced-3 mutant I248.	0.00e+00
39	3576	99.3	503	9	R45281	Ced-3 mutant K351.	0.00e+00
40	3576	99.3	503	9	R45312	Ced-3 mutant K323.	0.00e+00
41	3576	99.3	503	9	R45310	Ced-3 mutant K349.	0.00e+00
42	3576	99.3	503	9	R45310	Ced-3 mutant K321.	0.00e+00
43	3576	99.3	503	9	R45304	Ced-3 mutant L488.	0.00e+00
44	3575	99.2	503	9	R45326	Ced-3 mutant L381.	0.00e+00
45	3575	99.2	503	9	R45292	Ced-3 mutant L286.	0.00e+00

ALIGNMENTS

RESULT 1
 ID R47466 standard; Protein; 503 AA.
 AC R47466;
 DT 23-JUN-1994 (first entry)
 DE Ced-3.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophobic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 OS Caenorhabditis elegans.
 PN W09325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 WP1: 94-007542/01.
 DR N-PSDB: 054666.
 PR Isolated C elegans cell death genes ced-3 and ced-4 - used to develop agents to increase or prevent cell death in organisms
 PS Claim 6: Fig 4: 127bp; English.
 CC This sequence is encoded by the C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and was most abundant in embryos, but was also detected in larvae and young adults, suggesting that ced-3 is expressed not only in cells undergoing cell death. The four largest introns as well as sequences 5' of the start codon contain repetitive elements, some of which have been characterised in non-coding regions of other C. elegans genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is highly hydrophobic with no significant hydrophobic region that might be a transmembrane region. One region of Ced-3 is very rich in serine. It is thought that this region is involved in protein-protein interactions, similar to acid blobs in transcription factors. Of the mutations which occur within the ced-3 gene, eight of the mutations are missense mutations, two are nonsense mutations and two are putative splicing mutations (see also R53279-88). These mutations establish the null phenotype of the ced-3 gene, confirming that ced-3, like ced-4, function is not essential for viability. The ced-3 and ced-4 gene products may be used to develop agents for treating conditions

3.0%; Score 30; DB 16; Length 281;

```

"organism"="Arabidopsis thaliana"
"strain"="Columbia"
"/note"="vector: Belozagii; Site-1: HindIII; Site-2:
HindIII; Produced by Rod Wing"
/db_xref="taxon:3702"
/clone="T2F1"
/clone_11b="TAMU"

```

```

/organism="Arabidopsis thaliana"
/strain="Columbia"
/notes="vector: BelobACII; Site_1: HindIII; Site_2: HindIII; Produced by Rod Wing"
/db_xref="taxon:3702"
/clone="T17A8"

```

[illegible]

NID	92864900
VERSION	GI:2864900
KEYWORDS	884753.1
SOURCE	EST.
ORGANISM	<i>Dictyostellium discoideum</i> .
REFERENCE	<i>Dictyostellium discoideum</i> .
AUTHORS	Eukaryota; Dictyostelida. <i>Dictyostellium</i>
	1 (bases 1 to 560)
	Mizuno, H., Morio, T. and Tanaka, Y.

JOURNAL Unpublished (1998)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
 Department of Cyto genetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.astl.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
 University, Yongin, Korea. 449-728 bhnahebio@server.myongji.ac.kr
 Seq primer: M13 Reverse Primer.

FEATURES
 source
 Location/Qualifiers

1..247
 /organism="Oryza sativa"
 /cultivar="Milyang23"
 /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional cDNA library inserted into lambda ZAPII
 vector at 5' end with EcoRI and 3' end with Xho I site."
 /db_xref="taxon:4530"
 /map="6"
 /clone="97SN1784"
 /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
 /issue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /lab="16 c 21 g 34 t 169 others"
 BASE COUNT
 ORIGIN 7 a 16 c 21 g 34 t 169 others

Query Match 4.7%; Score 47; DB 17; Length 247;
 Best Local Similarity 14.0%; Pred. No. 1,40e-22;
 Matches 30; Conservative 97; Mismatches 86; Indels 2; Gaps 2;

18 AMANKHTHMTBMCVRRVGTNNKNGRTTNNDCSDNAHCXYTBWYARSKYGY 77
 1976 AAAAATTCAGAGAGGCTCAAAATTTCCGATTTACTGACTTCTTTTTCGTA 2035
 78 GTBYASNNVDINTGTGCTVNVHSGNNRCSNVVYVWBYACDHYHDBRANHDD 137
 2036 TCAGTCAGCTGTG-GAGTTTTCAGCAAACTAGGAAAAAATGATAAAAATTA 2094
 138 TCTNDRCYCNVTASDNGTSATKRVTYGDTSDCGG-CWRKVTYGSBYRCVNV 196
 2095 AATTCAGCTGATTTTGAAGACATTTTAAAAAACAATTTTCCATATATTTTC 2154
 197 RTSMATDKSTKBSMDMSRRSRVHYGRMBNKKR 231
 2155 ACTCATTTTCAGACTAATCAAAATCAATCATGTA 2189

RESULT 7
 LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
 DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
 ACCESSION AA754458
 NID 92801164
 VERSION AA754458.1 GI:2801164
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Oryza.
 1 (bases 1 to 247)
 Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
 Kim,W.T., Kim,W.T., Yang,W.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
 Lee,M.C. and Eun,M.Y.
 Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
 Unpublished (1998)
 On Jan 14, 1998 this sequence version replaced gi:1797455.
 CONTACT: Eun M.Y.

Department of Cyto genetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.astl.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
 University, Yongin, Korea. 449-728 bhnahebio@server.myongji.ac.kr
 Seq primer: M13 Reverse Primer.

FEATURES
 source
 Location/Qualifiers

1..247
 /organism="Oryza sativa"
 /cultivar="Milyang23"
 /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional cDNA library inserted into lambda ZAPII
 vector at 5' end with EcoRI and 3' end with Xho I site."
 /db_xref="taxon:4530"
 /map="6"
 /clone="97SN1784"
 /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
 /issue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /lab="16 c 21 g 34 t 169 others"
 BASE COUNT
 ORIGIN 7 a 16 c 21 g 34 t 169 others

Query Match 4.4%; Score 44; DB 17; Length 247;
 Best Local Similarity 11.9%; Pred. No. 2.20e-19;
 Matches 29; Conservative 112; Mismatches 99; Indels 3; Gaps 3;

5 TMNTVNRGCCBMMNKHMTBMCVRRVGTNNKNGRTTNNDCSDNAHCY 64
 2799 TCGCAAAATTCACAAATTTCCGCTTTCGCCAGCAAAATAGCTTAAAAAATAAT 2740
 65 TBYVYARSKYGY-GTBYASNNVDINTGTGCTVNVHSGNNRCSNVVYVWBYAC 123
 2739 CACTCGAGCTTTCAGTAATAATTAATTTCACTCAATAATTTGGGAAAGTCATA 2680
 124 DYHBYDBRANHDDTCTNDRCYCNVTASDNGTSATKRVTYGDTSDCGGCMRWY 183
 2679 TTCCGATTTTCGAGAGAAACCGTCGCGGTGCAAAATTCAGATCCGCCCTTT 2620
 184 SSBYRCGVNVVTTTSMATDKSTKBSMDMSR-RSRVHYGRMBNKKR 241
 2619 TTTCGAGAGACCTCTTTTGGCGCACTTAAACCTAGATCTGGCGAGAGAT 2560
 242 TWR 244
 2559 CAA 2557

RESULT 8
 LOCUS C70143 360 bp mRNA EST 29-SEP-1997
 DEFINITION C70143 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone yk3833 5', mRNA sequence.
 ACCESSION C70143
 NID 92440668
 VERSION C70143.1 GI:2440668
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilita;
 Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita;
 1 (bases 1 to 360)
 Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
 Sano,M., Miyata,A. and Nishigaki,A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 On Apr 14, 1993 this sequence version replaced gi:692895.
 CONTACT: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics

REFERENCE
AUTHORS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
1 (bases 1 to 252)
Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C., and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797457.

TITLE
JOURNAL
COMMENT

Contact: Eun M.Y.
Department of Cyto genetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

FEATURES
source

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN
Query Match 5.0%; Score 60; DB 17; Length 252;
Best Local Similarity 15.2%; Pred. No. 4.39e-37;
Matches 34; Conservative 110; Mismatches 76; Indels 4; Gaps 4;
Db 31 WCVASHGNCVSVNCTBRGTHCCKKVNWSTM-TWGTVMKBNVSGDMHVBNTKVDV 89
Cp 2349 TATACATTCACACATCTTTGCGATGACACTCGAGATTTCATCGACTTTAGAT 2290
Db 90 GNHTRCSWRBVTNMAHYDNTNCBRYNNNDYHMHMBMBTGCMTCTMCMWHNTKC 149
Cp 2289 GACTAGAGACATCATATGTTCTCTCTAGCAAGCTCCTCTATCT-TGACGCATCATC 2231
Db 150 TASGMHTSTNYDKSSTNTWGTBYSIDKSMHGWCSB-BVKYHTKVSITRATNSYTCVR 208
Cp 2230 TTGGCGTATGCTGACCTCTCTACTGACCGGATGACAGATGATGAT-TTGCAT 2172
Db 209 YCVAMMTKKVYKHYVVBGCHBTDSCKTMMNTNKHMTSTD 252
Cp 2171 TTAGTCGAAATGAGTGAATATTTGGAATAATAGTGTTTT 2128

RESULT 5
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)

AUTHORS

Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C., and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797457.

TITLE
JOURNAL
COMMENT

Contact: Eun M.Y.
Department of Cyto genetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

FEATURES
source

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN
Query Match 5.0%; Score 50; DB 17; Length 252;
Best Local Similarity 9.8%; Pred. No. 7.72e-26;
Matches 23; Conservative 121; Mismatches 87; Indels 4; Gaps 4;
Db 18 WMTTSBCHGNVWVCVASHGNCVSVNCTBRGTHCCKKVNWSTM-TWGTVMKBNVSGDM 77
Oy 2430 TCATGTGGAACGGTTCGCGAGAGAGACGGGAGATCG-TGAAGAGATGCAACACGGGG 2488
Db 78 HWYBANTVDVGNHTRCSWRBVTNMAHYDNTNCBRYNNNDYHMHMBMBTGCMTCT 137
Oy 2489 AGATGGGGCTTCGACGGCTTTATGATGCTCTGCTACGGGACGACGACTTGC 2548
Db 138 TMCMBHNTKCTASGMHTSTNYDKSSTNTWGTBYSIDKSMHGWCSB-BVKYHTKVSIT 197
Oy 2549 TGA-AGTCTCTGACCTCTCGCCAGATC-GTAGGTT-TTTAAAGTTTCGGCGCAAGCAA 2605
Db 198 RATRSTCYRKCYVMWMTKKVYKHYVVBGCHBTDSCKTMMNTNKHMTSTD 252
Oy 2606 GGCTCTCAGCAAAAGAGGCGGATCGTAATTTGCAACCCACCGGCGCGTT 2660

RESULT 6
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754458
NID 92801164
VERSION AA754458.1 GI:2801164
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 247)
Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C., and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

RESULT	3	
LOCUS	C51090	377 bp mRNA EST 11-SEP-1997
DEFINITION	C51090 yuji kohara unpublished cdna Caenorhabditis elegans cdna clone yk491a5 5', mRNA sequence.	
ACCESSION	C51090	
NID	92388343	
VERSION	C51090.1	GI:2388343
KEYWORDS	EST.	
SOURCE	Caenorhabditis elegans.	
ORGANISM	Caenorhabditis elegans.	
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditidae; Peloderiinae; Caenorhabditis	
AUTHORS	1 (bases 1 to 377) Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.	
TITLE	Expression map of the C.elegans genome	
JOURNAL	Unpublished (1996)	
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1405104.	
FEATURES		
source	<p>Contact: Yuji Kohara Gene Library Lab National Institute of Genetics Yata 111, Mishima, Shizuoka 411, Japan Tel: 0559-75-0771 Fax: 0559-75-6240 Email: ykohara@dbj.nig.ac.jp. Location/Qualifiers 1..377 /organism="Caenorhabditis elegans" /strain="CB1489 him-8(e1489)" /note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal" /db_xref="taxon:6239" /map="11q23" /clone="yk491a5" /clone_11b="Yuji Kohara unpublished CDNA" /clone_11b="74 c 103 g 92 t 2 others</p>	
BASE COUNT	106 a 74 c 103 g 92 t 2 others	
ORIGIN		
Query Match	16 2%; Score 162; DB 15; Length 377;	
Best Local Similarity	98.8%; Pred. No. 1,19e-165;	
Matches	168; Conservative 0; Mismatches 1; Indels 1; Gaps 1;	
Db	1 TACNGTCATCTAAGAG-TCACACATCACCCCAACATGATCGCTCAAGATGAGAGAGCT 59	
Y	2197 TACGGCGCATTAAGAGGCTCAACCATCACACCCCAACATGATCGCTCAAGATGAGAGAGCT 2256	
Db	60 TGTAGAGAGAACATTTATGATGTTCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAG 119	
Y	2257 TGTAGAGAGAACATTTATGATGTTCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAG 2316	
Db	120 TTTCATCCGCAAAACAAGTGTGAATAGAGATAATGAGAGATATGATTAAAT 169	
Y	2317 TTTCATCCGCAAAACAAGTGTGAATAGAGATAATGAGAGATATGATTAAAT 2366	
RESULT	4	
LOCUS	AA754459	252 bp mRNA EST 20-JAN-1998
DEFINITION	97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa	
ACCESSION	AA754459	
NID	92801165	
VERSION	AA754459.1	GI:2801165
KEYWORDS	EST.	
SOURCE	Oryza sativa.	
ORGANISM	Oryza sativa	

 WIRE (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_mn n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Fri Aug 6 03:42:06 1999; MasPar time 1720.16 Seconds
 Tabular output not generated.

Title: >US-08-287-669-18
 Description: (1801-2800) from US08287669.seq (3 of 10)
 Perfect Score: 1000
 N.A. Sequence: 1801 TTCGAGCTGACAAACAGAA.....CATTTGTGATTTGCCGAA 2800
 Comp: AAGGTCGACACTGTTCCTT.....GTAAACATCTAAACGGCTT

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb1-est158
 1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est12
 6:em_est9 7:em_gss1
 Database: genbank-est111
 8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
 13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
 17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
 21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
 25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
 29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
 34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_est39
 39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.717; Variance 3.590; scale 3.264
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB	ID	Description	Pred. No.
1	198	19.8	360 32	D75533	CEIK105H1F yuji Kohara	1.95e-213
2	177	17.7	360 15	C39290	yuji Kohara unip	1.76e-185
3	162	16.2	377 15	C51090	yuji Kohara unip	1.19e-165
4	60	6.0	252 17	AA754459	97SN1787 Rice	4.39e-37
5	50	5.0	252 17	AA754459	97SN1787 Rice	7.72e-26
6	47	4.7	247 17	AA754458	97SN1784 Rice	1.40e-22
7	44	4.4	247 17	AA754458	97SN1784 Rice	2.20e-19
8	36	3.6	360 15	C70143	yuji Kohara unip	2.95e-11
9	36	3.6	2275 20	AF034173	Human mRNA (T	2.95e-11
10	34	3.4	2275 20	AF034173	Human mRNA (T	2.46e-09

Result No.	Score	Match Length	DB	ID	Description	Pred. No.
11	31	3.1	580 18	C84753	C84753 Dictyostelium d	1.48e-06
12	31	3.1	779 37	B21852	T2F1-Sp6 TMU Arabidop	1.48e-06
13	31	3.1	876 37	B12885	T17A8-T7 TMU Arabidop	1.48e-06
14	30	3.0	281 16	AA633064	ng10e07.s1 NCI.CGAP.T	1.16e-05
15	30	3.0	401 42	AA0445355	GSTC01014 Trypanosoma	1.16e-05
16	30	3.0	503 37	B08074	TG24-Sp6.1 TMU Arabi	1.16e-05
17	30	3.0	727 37	B12502	F18F5-Sp6.1 IGF Arabid	1.16e-05
18	30	3.0	809 37	B10155	F4F1-Sp6.1 IGF Arabid	1.16e-05
19	29	2.9	161 24	AI204411	qf72e04.x1 Soares-test	8.77e-05
20	29	2.9	187 8	Z24996	HSB82C032 STRATAGENE H	8.77e-05
21	29	2.9	215 26	AI381412	lc50f01.x1 Soares-test	8.77e-05
22	29	2.9	231 19	R25172	yh37c06.s1 Soares-plac	8.77e-05
23	29	2.9	244 24	AI198459	gf88a07.x1 Soares-plac	8.77e-05
24	29	2.9	300 21	AA953741	oc37c09.s1 NCI.CGAP.Lu	8.77e-05
25	29	2.9	308 35	AA039740	z10g10.s1 Soares-feta	8.77e-05
26	29	2.9	323 23	AI147215	q89c09.x1 Soares-feta	8.77e-05
27	29	2.9	335 8	T18123	0677c3 czapFD02.1, De	8.77e-05
28	29	2.9	370 15	AA593600	no26d09.s1 NCI.CGAP.Pr	8.77e-05
29	29	2.9	374 18	AA783005	a158a06.s1 Soares-test	8.77e-05
30	29	2.9	374 20	AA918212	on82a01.s1 Soares-NFL	8.77e-05
31	29	2.9	377 30	R62771	y110i10.s1 Soares-plac	8.77e-05
32	29	2.9	388 24	AI220889	q903c07.x1 Soares-plac	8.77e-05
33	29	2.9	424 16	AA622274	no42h08.s1 NCI.CGAP.Pr	8.77e-05
34	29	2.9	440 20	AA907140	cd44c05.s1 NCI.CGAP.GC	8.77e-05
35	29	2.9	441 23	AI138427	q84a407.x1 Soares-test	8.77e-05
36	29	2.9	444 23	AI138414	q84a404.x1 Soares-test	8.77e-05
37	29	2.9	456 32	N26720	yx61i09.s1 Soares-mela	8.77e-05
38	29	2.9	494 35	AA039776	z110c10.s1 Soares-feta	8.77e-05
39	29	2.9	505 29	AA568541	tn40e10.x1 NCI.CGAP.Br	8.77e-05
40	29	2.9	505 17	AA722943	z985a06.s1 Soares-feta	8.77e-05
41	29	2.9	505 24	AI188773	q616f05.x1 Soares-plac	8.77e-05
42	29	2.9	535 35	AA045000	zk63a04.s1 Soares-plac	8.77e-05
43	29	2.9	541 30	H09727	y190e11.s1 Soares-feta	8.77e-05
44	29	2.9	601 15	AA573749	nk07b07.s1 NCI.CGAP.Co	8.77e-05
45	29	2.9	844 37	B10796	T26g15-Sp6 TMU Arabid	8.77e-05

ALIGNMENTS

RESULT 1
 LOCUS D75533 360 bp mRNA
 DEFINITION CEIK105H1F yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk105h1 5', mRNA sequence.
 ACCESSION D75533
 NID 91121320
 VERSION D75533.1 GI:1121320
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 360)
 Kohara, Y., Mitsukuri, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
 Tabara, H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT

CONTACT: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykoha@dbj.nig.ac.j
 Insert length: 2279 Std Error: 0.00.
 location/Qualifiers

FEATURES
 source
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /note="dev stage=varied, sex=Hermaphrodite male,
 tissue=whole animal"
 /db_xref="taxon:6239"
 /clone="yk105h1"

FT	/rpt_type= INVERTED
FT	/note="Repeat 3"
FT	5221..5330
FT	repeat_unit
FT	/*tag= q
FT	/rpt_type= INVERTED
FT	/note="Repeat 3"
FT	5547..5760
FT	/*tag= r
FT	/number= Exon_5
FT	5761..5814
FT	/*tag= s
FT	/number= Intron_5
FT	5815..5942
FT	/*tag= t
FT	/number= Exon_6
FT	5943..6297
FT	/*tag= u
FT	/number= Intron_6
FT	6062..6138
FT	/*tag= v
FT	/rpt_type= INVERTED
FT	6298..6537
FT	/*tag= w
FT	/number= Exon_7
FT	6538..7012
FT	/*tag= x
FT	/number= Intron_7
FT	6567..6625
FT	/*tag= y
FT	/rpt_type= INVERTED
FT	/note="Inverted w.r.t. repeat at 6905-6965"
FT	6905..6965
FT	repeat_unit
FT	/*tag= z
FT	/rpt_type= INVERTED
FT	/note="Inverted w.r.t. repeat 6567-6625"
FT	7012..7075
FT	/*tag= aa
FT	/number= Exon_8
PN	MO9325685-A.
PD	23-DEC-1993.
PF	14-JUN-1993; 005701.
PR	12-JUN-1992; US-897788.
PR	20-NOV-1992; US-979638.
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.
PI	Horvitz HR, Shahan S, Yuan J;
PI	WPI; 94-007542/01.
DR	P-PSDS; R53380.
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT	develop agents to increase or prevent cell death in organisms
PS	Claim 14; Fig 4; 127BP; English.
CC	The sequences given in 064735-45 represent mutations of the C. elegans
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC	was most abundant in embryos, but was also detected in larvae and young
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing
CC	cell death. The four largest introns as well as sequences 5' of the
CC	start codon contain repetitive elements, some of which have been
CC	characterised in non-coding regions of other C. elegans genes, such
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC	length. Ced-3 is highly hydrophobic with no significant hydrophobic
CC	region that might be a transmembrane region. One region of Ced-3 is
CC	very rich in serine. It is thought that this region is involved in
CC	protein-protein interactions, similar to acid blobs in transcription
CC	factors. Of the mutations which occur within the ced-3 gene, eight of
CC	the mutations are missense mutations, two are nonsense mutations and
CC	two are putative splicing mutations. These mutations establish the
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC	function is not essential for viability. The ced-3 and ced-4 gene
CC	characters may be used to develop agents for treating conditions
CC	characterised by cell deaths, such as myocardial infarction, stroke,
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC	infection, aging or hair loss.
CC	Sequence 7653 BP: 2430 A; 1453 C; 1271 G; 2499 T;

Query Match	98.8%	Score 988	DB 9	Length 7653
Best Local Similarity	99.4%	Pred. No. 0.00e+00		
Matches 994	Conservative 0	Mismatches 6	Indels 0	Gaps 0
Db	1801	ttccagagctgcgaacacagaaacaaacacacacacacacacattttaaaatcgatttcaat	1860	
Qy	1801	ttccagagctgcgaacacagaaacaaacacacacacacacacattttaaaatcgatttcaat	1860	
Db	1861	taaaaataacgatttctcatgtgaatatgtgttttatagttgcgaataaagaagac	1920	
Qy	1861	taaaaataacgatttctcatgtgaatatgtgttttatagttgcgaataaagaagac	1920	
Db	1921	gattcaaaaacatttlaacaaaaaaaccccaaatcgcgagaaatcaagataaaaa	1980	
Qy	1921	gattcaaaaacatttlaacaaaaaaaccccaaatcgcgagaaatcaagataaaaa	1980	
Db	1981	ttcaagaggggtcaaatcttcacgatttaccgacttcaacttttttgcgtgaagt	2040	
Qy	1981	ttcaagaggggtcaaatcttcacgatttaccgacttcaacttttttgcgtgaagt	2040	
Db	2041	gcagctgtctggagttttttgacgaaacactgagaaataacgatataaaattactcaatcg	2100	
Qy	2041	gcagctgtctggagttttttgacgaaacactgagaaataacgatataaaattactcaatcg	2100	
Db	2101	agctgaattttgaggacaatgtttaaaaaaacactatttttccaataattcaactcat	2160	
Qy	2101	agctgaattttgaggacaatgtttaaaaaaacactatttttccaataattcaactcat	2160	
Db	2161	tttcgactaataatcgaaatacaaatcgtaacttcaactgaggtctgaagaggtcaacc	2220	
Qy	2161	tttcgactaataatcgaaatacaaatcgtaacttcaactgaggtctgaagaggtcaacc	2220	
Db	2221	atcaacccgaagaatgatgctcaagaataagaagagcttgcctcagaagagaacattatgtt	2280	
Qy	2221	atcaacccgaagaatgatgctcaagaataagaagagcttgcctcagaagagaacattatgtt	2280	
Db	2281	ttctcagtaacttcaaaagtcgaatattctcgaagtttccatcgcgaacaagaatgttga	2340	
Qy	2281	ttctcagtaacttcaaaagtcgaatattctcgaagtttccatcgcgaacaagaatgttga	2340	
Db	2341	atagatgaataatggagagatagttaatgttgagttttaaactcgataataattttaaaaaa	2400	
Qy	2341	atagatgaataatggagagatagttaatgttgagttttaaactcgataataattttaaaaaa	2400	
Db	2401	aattgataataaagaatatattttgcagtaatgttgaaacggttcgcgagaagaagacggg	2460	
Qy	2401	aattgataataataaagaatatattttgcagtaatgttgaaacggttcgcgagaagaagacggg	2460	
Db	2461	agatcgtgaagaacgctgcaacgacgagagagatgtgagcttccacgcttttatagtctc	2520	
Qy	2461	agatcgtgaagaacgctgcaacgacgagagagatgtgagcttccacgcttttatagtctc	2520	
Db	2521	ttcgcctcaaggagacagaaagacttgcgaagtcttctgaagctctcgcgaagatcgtagg	2580	
Qy	2521	ttcgcctcaaggagacagaaagacttgcgaagtcttctgaagctctcgcgaagatcgtagg	2580	
Db	2581	tttttaaaagttccggcgcaaaagcgtctcagcaggaataaagaagcgcgaatcgttaattt	2640	
Qy	2581	tttttaaaagttccggcgcaaaagcgtctcagcaggaataaagaagcgcgaatcgttaattt	2640	
Db	2641	gcaacccacgcgcacggtttttctcctcgaaaaatcgyaaaatlabgcacttcccaatat	2700	
Qy	2641	gcaacccacgcgcacggtttttctcctcgaaaaatcgyaaaatlabgcacttcccaatat	2700	
Db	2701	tttgaagtgaataatatttatttattcctcgaagcctcgagtgattattttttttaaacta	2760	
Qy	2701	tttgaagtgaataatatttatttattcctcgaagcctcgagtgattattttttttaaacta	2760	
Db	2761	atttcgtgcgcaaaagcgcattttgagatttgcgaa	2800	
Qy	2761	atttcgtgcgcaaaagcgcattttgagatttgcgaa	2800	

	CC	cell death. The four largest introns as well as sequences 5' of the
	CC	start codon contain repetitive elements, some of which have been
	CC	characterised in non-coding regions of other C. elegans genes, such
	CC	as Iem-1, lin-12 and myod. The Ced-3 protein is 503 amino acids in
	CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic
	CC	region that might be a transmembrane region. One region of Ced-3 is
	CC	very rich in serine. It is thought that this region is involved in
	CC	protein-protein interactions, similar to acid blobs in transcription
	CC	factors. Of the mutations which occur within the ced-3 gene, eight of
	CC	the mutations are missense mutations, two are nonsense mutations and
	CC	two are putative splicing mutations. These mutations establish the
	CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
	CC	function is not essential for viability. The ced-3 and ced-4 gene
	CC	products may be used to develop agents for treating conditions
	CC	characterised by cell deaths, such as myocardial infarction, stroke,
	CC	infectious disease, traumatic brain injury, hypoxia, pathogenic
	CC	infection, aging or hair loss.
SQ	Sequence	7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
	Query Match	99.0%; Score 990; DB 9; Length 7653;
	Best Local Similarity	99.5%; Pred. No. 0.0ne+00;
Matches	995; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
Db	1801	tcccgagctgcacaaacgagaacaacaacaacatttaaaaaatcagtttccaat 1860
OY	1801	tttcaggctgcgaamaacagaaaacacaaacacataattttaaaatcagtttccaat 1860
Db	1861	taaaaatacagattcttcattganaaatgttgttttagttggccaaaataaagagact 1920
OY	1861	TAAAAATACGATTCTCATTTGAAAATTTGTATTAGTTCGCAAAATAAAGAGACT 1920
Db	1921	gatccaacaatttaaacaacaaaccccccaaatcgcgcgagaaatcaagataaaaaa 1980
OY	1921	GATTCAAAACAATTTTAACAAAAAAACCACAAATTCGCCGAAMTCAAGTAFAAAA 1980
Db	1981	ttcaaagaggttcaaaattttcgcattttacctgttccacttttttttgtagtcagt 2040
OY	1981	TTCAAAGGGTGTAATAATTTCCGATTTTACGACTTTTCACCTTTTTTTTGTAGTTCAGT 2040
Db	2041	gcagttgtttgagttttttgcagaaactagaaaataacgataaaattaccacaatcg 2100
OY	2041	GCAGTTGTGGAGTTTGTGACGAAACCTAGSAAAAAATGATRAAAATTTACTCAANTCG 2100
Db	2101	agctgaatttggaggacaaigtcttaaaaaaacataattttccaataattcacatcat 2160
OY	2101	AGCTGAATTTTGAGACACATGTTAAAAAAAACACATAATTTCCAATATTTCACCTCAT 2160
Db	2161	tttcagactaaatcgaaatacaatctgatatcttgcagtaagggctcagtagagagttcaacc 2220
OY	2161	TTTCGACTTAATCGMAAATTAATTCGTCTTGACTACGGGTGACGTAGAGAGGTCAACC 2220
Db	2221	atcagccgaagaatgatcgctaaagataagaagagctcttcagagaggaacattatgct 2280
OY	2221	ATCAGCCGAAGATGATCGCTCAACATATGAGAAGAGCTTGCTAGAGAGCACTATTATGAT 2280
Db	2281	tctcgaagcactctaaggtcogatgaaattctcgaagttctcatcgcaaaaagaigtga 2340
OY	2281	TCTCTAGTCATCTAAAGGTGATGAATTTCTCCAACTTCTCAACGCCAAAAACAAGTGTGA 2340
Db	2341	aatagtgataatcagagatacatatcaatcagtttccaatcgaaataaattttaaaaaa 2400
OY	2341	AATAGGTAATAGGAGATATGATTAAATGATGAGTTTAAATCGAATTAATTTAAAAANA 2400
Db	2401	aattgataataataaagaatatlttcgcagtcatalcvgnaacygtlctcgcgaaagagacgg 2460
OY	2401	AATTGATAATAATAAAGATAATTTTTCGAGTCATGTGAACGGTTCCGCGAGAGACGGG 2460
Db	2461	agatcgtgaaagcagctgcacgcgagggagatgctgcgttcaacgcgttttatgcttc 2520
OY	2461	AGATCGTGAAGACGATGCACGACGGGAGATGTGGCGTTTCACGCGTTTATGATGCTC 2520
Db	2521	ttcgcttacggacacgaagagcttgcgaagtttcttgaagctctcgcgcaatcgttagg 2580

FT		/*tag= aa
FN	/number= Exon_8	
PT	W09325685-A.	
PD	23-DEC-1993.	
PF	14-JUN-1993: U05701.	
PR	12-JUN-1992: US-897788.	
PA	20-NOV-1992: US-89638.	
PB	(MAST) MASSACHUSETTS INST TECHNOLOGY.	
PI	Horvitz HR, Shaham S, Yuan J;	
DR	MP1; 94-007542/01.	
DR	P-PDB: R53283.	
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to	
PT	develop agents to increase or prevent cell death in organisms	
PS	Claim 14, Fig 4: 127bp, English.	
CC	The sequences given in Q64735-45 represent mutations of the C. elegans	
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and	
CC	was most abundant in embryos, but was also detected in larvae and young	
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing	
CC	cell death. The four largest introns as well as sequences 5' of the	
CC	start codon contain repetitive elements, some of which have been	
CC	characterised in non-coding regions of other C. elegans genes, such	
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in	
CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic	
CC	region that might be a transmembrane region. One region of Ced-3 is	
CC	very rich in serine. It is thought that this region is involved in	
CC	protein-protein interactions, similar to acid blobs in transcription	
CC	factors. Of the mutations which occur within the ced-3 gene, eight of	
CC	the mutations are missense mutations, two are nonsense mutations and	
CC	two are putative splicing mutations. These mutations establish the	
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,	
CC	function is not essential for viability. The ced-3 and ced-4 gene	
CC	products may be used to develop agents for treating conditions	
CC	characterised by cell deaths, such as myocardial infarction, stroke,	
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic	
CC	infection, aging or hair loss.	
SQ	Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;	
Query Match	99.0%; Score 990; DB 9; Length 7653;	
Best Local Similarity	99.5%; Pred. NO. 0.00e+00;	
Matches	995; Conservative 0; Mismatches 5; Indels 0; Gaps 0	
Dn	1801 ttccagctgcaaaaacagaaacaaaaaacacacaacaattttaaaatcagtlttccaat	1860
Oy	1801 TTCCAGCTGACAAACAACAAAACAAAAACACACAAATTTTAAATCAGTTTCAAT	1860
Dn	1861 taaaataacagtttcattcatgaattgttgttttaagtttgcccaaatataaagaagact	1920
Oy	1861 TAAATAATACGATTTCATTGAATTTGTGTTTAATGTTTGGAATAAAGAGACT	1920
Dn	1921 gatccaacaacattttaaacaataaaaccoccaaattgcgcgaatacaagataaaaa	1980
Oy	1921 GATTCAAACCATTTTAAACAAAAAACCCCAAAATGCCCGAGAATCAATATAAAAA	1980
Dn	1981 ttcaagaagggtcaaaaattttccgatttctaactgacttacaccttttttcgtagttcagt	2040
Oy	1981 TTCAAAGGGGTAAAAATTTCCGATTTTACTGCATTTTCACCTTTTTTTTCGTAGTTCAGT	2040
Dn	2041 ggcagttgttgagagtttttggacgaaactagaaaaaaatcgataaaatatctccaatcg	2100
Oy	2041 GCAGTTGTGGAGTTTGGACGAAAAC TAGAAAAAATCGATAAAATTTCTCAAAATCG	2100
Dn	2101 agctgaattttagagcaaatgttttaaaaaaaacactattttccataatttcactat	2160
Oy	2101 AGCTGAATTTTAGAGCAATGTTTAAAAAAAACACATTTTCCATATTTCACTCAT	2160
Dn	2161 ttccaacaaatcgaaaataaataatcgtactcgtactagcgtccagttagagaggtccaac	2220
Oy	2161 TTTCACAACTAAATCGAAAATCAATTCGACTCTGACTACGGGTCAGTAAGAGGTCAACC	2220
Dn	2221 atcaagccgaagatctgctcaataatagaagaagctgtcttagagagaaacattatgagt	2280
Oy	2221 ATCAGCCGAAGATGATGCTCAAGATAGAAAGAGGAGCTTGCTAAGAGAGAAATATGAGT	2280

```

FT repeat_unit 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT 7012..7075
FT /*tag= aa
FT /number= Exon_8
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR WPI: 94-007542/01.
P-PSDB: R53286.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127bp; English.
CC The sequences given in 064735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterized in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterized by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.0%; Score 990; DB 9; Length 7653;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Db 2161 ttccagactaaatcgaaatcaaatctgactgactagcgtcagtagagagttcaacc 2220
Qy 2161 TTTTCAGACTAAATCGAAATCAATCTGACTAGCGGTCACTAGAGAGTCAACC 2220
Db 2221 atcagccgaagatgatgcgtcgaagatagaagagctctgtagagagaaatgatgt 2280
Qy 2221 ATCAGCCGAAGATGATGCGTCAAGATAGAAGAGCTTGTGAGAGAGAACATATGATGT 2280
Db 2281 tctctagtcattcaaaagtcgatgaaattctcgaaattctcatcgcaaaacaaagtgtga 2340
Qy 2281 TCTCTAGTCATTCAAAAGTCGATGAAATCTCGAAATCTCAATCGCAAAACAAAGTGTGA 2340
Db 2341 atagtgtaattgagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2400
Qy 2341 ATAAGTGAATTAAGAGATATATATATATATATATATATATATATATATATATATATATAT 2400
Db 2401 aatgataataataaagaataatttgcagtcattgtaacagcttcgagaaagagagcg 2460
Qy 2401 AATGATAATATAAAGAAATTTTTCAGTCATGTGGAACGGTTCGAGAAAGAGACGGG 2460
Db 2461 agatcgtgaagaagcagtcgaacgacgagggagagatggtcgtccacggtttatagatc 2520
Qy 2461 AGATCGTGAAGACAGTCACACGAGGAGATGTGCGGTTCGACGGCTTATGATGCTC 2520
Db 2521 ttgcgtctacggagcagcagagagcttgtaagttcttgaagcttcgacagatcgtag 2580
Qy 2521 TTGCTCTACGGAGACGAGAGAGCTTCTAATTTCTTGAACTCTCGCAGATCTGATAG 2580
Db 2581 tttaaagttcgcgcgaagaagaaggttcacggaataaagaagcgatcgttaatt 2640
Qy 2581 TTTTAAAGTTGCGCGCAAAAGCAAGGCTCTACGGAATAAAGAGCGATCGTAAATTT 2640
Db 2641 gcaaccacccgagcaggttttctcctcgaaatcggaattatgacattcccaatt 2700
Qy 2641 GCAACCCACCGCGCAGGCTTTTCTCGAAATCGAAATATGCACTTCCCAATAT 2700
Db 2701 ttgaagtgaaatattatttactgaagctcgagtgatatttattttaaacta 2760
Qy 2701 TTGAAGTGAATATATTTTATTTACTGAAGCTCGAGTGTATTTATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaagcattttagatttgcgaa 2800
Qy 2761 ATTTCTGCGGCGCAAAAGGCCATTTGTAGATTGCCGA 2800

RESULT 9
ID 064740 standard; DNA; 7653 BP.
AC 064740:
DN 23-JUN-1994 (first entry)
DE ced-3 (C6322N) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH key location/Qualifiers
FH repeat_unit 1356..1472
FH /*tag= a
FH /rpt_type= INVERTED
FH /note= "Inverted w.r.t. repeat at 1490-1614"
FH 1490..1614
FH /*tag= b
FH /rpt_type= INVERTED
FH /note= "Inverted w.r.t. repeat at 1356-1472"
FH 2232..2366
FH /*tag= c
FH /number= Exon_1
FH 2367..2429
FH /*tag= d
FH /number= Intron_1
FH 2450..2575
FH /*tag= e
FH exon

```



```

OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /tag= g
FT /number= Exon_3
FT intron 3108..4302
FT /tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /tag= m
FT /number= Exon_4
FT intron 4635..5546
FT /tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT repeat_region 6062..6138
FT /tag= u
FT /rpt_type= INVERTED

```

```

FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT mutation 6535
FT /tag= w
FT /note= "G>A, from allele n2426"
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT W09325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993: U05701.
FT 12-JUN-1992: US-897788.
FT 20-NOV-1992: US-979638.
FT (MAST ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shahan S, Yuan J;
FT MPI: 94-007542/01.
FT P-PSDB: R53287.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 12/PP; English.
FT The sequences given in 064735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT start codon contain repetitive elements, some of which have been
FT characterised in non-coding regions of other C. elegans genes, such
FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT region that might be a transmembrane region. One region of Ced-3 is
FT very rich in serine. It is thought that this region is involved in
FT protein-protein interactions, similar to acid blobs in transcription
FT factors. Of the mutations which occur within the ced-3 gene, eight of
FT the mutations are missense mutations, two are nonsense mutations and
FT two are putative splicing mutations. These mutations establish the
FT null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT function is not essential for viability. The ced-3 and ced-4 gene
FT products may be used to develop agents for treating conditions
FT characterised by cell deaths, such as myocardial infarction, stroke,
FT degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT infection, aging or hair loss.
FT Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
SQ
Query Match 99.0%; Score 90; DB 9; Length 7653;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 1801 ttccagctgcgaacaaacaaacaaacaaacaaacatttaaaatcagtttcaaat 1860
Qy 1801 TTCAGCTGCACAAACGAACAAACAAACAAACAAACATTAAAAATCAGTTTCAAT 1860
Db 1861 taaaaatacgaattctcatgaaattcgttatgttgcgaataaagaagaact 1920
Qy 1861 TAAAAATAAGATTTCATTCGAAATTTGTTTATGTTTCGAAAAAAGAGAACT 1920
Db 1921 gattcaaaacaatttaacaaaaaaaccccaaaattgcgcagaataaagataaaaa 1980
Qy 1921 GATTCAAAACAATTTAACAAAAAAACCCCAAAATTCGCAGAAATCAAGATAAAAA 1980
Db 1981 ttcaagaggtcaaaatttcgatttactgacttacccttttttcgagttcagt 2040
Qy 1981 ttcaagaggtcaaaatttcgatttactgacttacccttttttcgagttcagt 2040

```


RESULT	6	standard; DNA: 7653 BP.
ID	064738	
AC	23-JUN-1994	(first entry)
DT	ced-3 (C5940T) gene.	
DE	ced-3 (C5940T) gene.	
CM	C. elegans; ced-4; ced-3; mutant; transcriptional regulation;	
KW	embryogenesis; cell death; hydrophobic; transmembrane; region;	
KW	hydrophobic; mutation; amino acid; substitution; RNA splicing	
KW	protein synthesis; null phenotype; calcium-binding domain; ss	
OS	Caenorhabditis elegans.	
FH	Key	Location/Qualifiers
FT	repeat_unit	1356..1472
FT		/*tag= a
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t repeat at 1490-1614"
FT	repeat_unit	1490..1614
FT		/*tag= b
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 1356-1472"
FT	exon	2232..2366
FT		/*tag= c
FT		/number= Exon_1
FT	intron	2367..2429
FT		/*tag= d
FT		/number= Intron_1
FT	exon	2450..2575
FT		/*tag= e
FT		/number= Exon_2
FT	intron	2576..2853
FT		/*tag= f
FT		/number= Intron_2
FT	exon	2854..3107
FT		/*tag= g
FT		/number= Exon_3
FT	intron	3108..4302
FT		/*tag= h
FT		/number= Intron_3
FT	repeat_unit	3126..3243
FT		/*tag= i
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3329-3396"
FT	repeat_unit	3329..3396
FT		/*tag= j
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3126-3243"
FT	repeat_unit	3487..3759
FT		/*tag= k
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3782-4070"
FT	repeat_unit	3782..4070
FT		/*tag= l
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat 3487-3759"
FT	exon	4303..4634
FT		/*tag= m
FT		/number= Exon_4
FT	intron	4635..5546
FT		/*tag= n
FT		/number= Intron_4
FT	repeat_unit	4688..4719
FT		/*tag= o
FT		/rpt_type= INVERTED
FT		/note= "Repeat 3"
FT	repeat_unit	5221..5330
FT		/*tag= p
FT		/rpt_type= INVERTED
FT		/note= "Repeat 3"
FT	exon	5547..5760
FT		/*tag= q
FT		/number= Exon_5
FT	intron	5761..5814
FT		/*tag= r
FT		/number= Intron_5

Db 2701 ttgaagtaataatatttactgaagctcgagatattatattttaaacta 2760
 QY 2701 TTGAAGTAATAATATTACTGAAGCTCGAGATATTATTTTAACTA 2760
 Db 2761 atttcgtggcgcaaaagccattttagatttgcgaa 2800
 QY 2761 ATTTCGTGGCGCAAAAGCCATTTTGTAGATTGGCGAA 2800

RESULT 5
 ID 064741 standard; DNA: 7653 BP.
 AC 064741;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (G6372A) gene.
 KM C. elegans; ced-3; mutant; transcriptional regulation;
 KM embryogenesis; cell death; hydrophilic; transmembrane; region;
 KM hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT repeat_unit 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1490-1614"
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT /tag= c
 FT /number= Exon_1
 FT /tag= d
 FT /number= Exon_1
 FT /tag= e
 FT /number= Exon_2
 FT /tag= f
 FT /number= Intron_2
 FT /tag= g
 FT /number= Exon_3
 FT /tag= h
 FT /number= Intron_3
 FT /tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT /tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT /tag= m
 FT /number= Exon_4
 FT /tag= n
 FT /number= Intron_4
 FT /tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT repeat_unit 5221..5330

FT /tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT /tag= q
 FT /number= Exon_5
 FT /tag= r
 FT /number= Intron_5
 FT /tag= s
 FT /number= Exon_6
 FT /tag= t
 FT /number= Intron_6
 FT /tag= u
 FT /rpt_type= INVERTED
 FT /tag= v
 FT /number= Exon_7
 FT /tag= w
 FT /note= "G>A, from allele n1286"
 FT /tag= x
 FT /number= Intron_7
 FT /tag= y
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT /tag= z
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT /tag= aa
 FT /number= Exon_8

WO9325685-A.
 PD 23-DEC-1993.
 PR 14-JUN-1993; US-05701.
 PR 12-JUN-1992; US-897768.
 PR 20-NOV-1992; US-979638.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI; 94-007542/01.
 DR P-PSDB: R53284.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 14; Fig 4; 12pp; English.
 CC The sequences given in 064735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 CC Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Accession	Gene	Protein	Protein Description
FT	W09325685-A.	23-DEC-1993.	number=Exon_4
FT	14-JUN-1993;	U05701.	/*tag= n
FT	12-JUN-1992;	US-897788.	number=Intron_4
FT	20-NOV-1992;	US-979638.	4688..4719
FT	(MASI) MASSACHUSETTS INST TECHNOLOGY.		/*tag= o
FT	PI Horvitz HR, Shaham S, Yuan J;	WPI: 94-007542/01.	/rpl_type= INVERTED
FT	p-RSDS; R53285.		/note="Repeat 3"
FT	Isolated C elegans cell death genes ced-3 and ced-4 - used to		5221..5330
FT	develop agents to increase or prevent cell death in organisms		/*tag= p
FT	PS Claim 14; Fig 4; 127pp; English.		/rpl_type= INVERTED
FT	The sequences given in 064735-45 represent mutations of the C. elegans		/note="Repeat 3"
FT	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and		5547..5760
FT	was most abundant in embryos, but was also detected in larvae and young		/*tag= q
FT	adults, suggesting that ced-3 is expressed not only in cells undergoing		number=Exon_5
FT	cell death. The four largest introns as well as sequences 5' of the		5761..5814
FT	start codon contain repetitive elements, some of which have been		/*tag= r
FT	characterised in non-coding regions of other C. elegans genes, such		number=Intron_5
FT	as fem-1, lin-12 and myoD. The ced-3 protein is 503 amino acids in		5815..5942
FT	length. Ced-3 is highly hydrophilic with no significant hydrophobic		/*tag= s
FT	region that might be a transmembrane region. One region of Ced-3 is		number=Exon_6
FT	very rich in serine. It is thought that this region is involved in		5943..6297
FT	protein-protein interactions, similar to acid blobs in transcription		/*tag= t
FT	factors. Of the mutations which occur within the ced-3 gene, eight of		number=Intron_6
FT	the mutations are missense mutations, two are nonsense mutations and		6062..6138
FT			/*tag= u
FT			/rpl_type= INVERTED
FT			6298..6537
FT			/*tag= v
FT			number=Exon_7
FT			6434
FT			/*tag= w
FT			/note="C>T, from allele n1129 and n1164"
FT			6538..7012
FT			/*tag= x
FT			number=Intron_7
FT			6567..6625
FT			/*tag= y
FT			/rpl_type= INVERTED
FT			/note="Inverted w.r.t. repeat at 6905-6965"
FT			6905..6965
FT			/*tag= z
FT			/rpl_type= INVERTED
FT			/note="Inverted w.r.t. repeat 6567-6625"
FT			7012..7015
FT			/*tag= aa
FT			number=Exon_8

Query Match	99.0%;	Score 990;	DB 9;	Length 7653;
Best Local Similarity	99.5%;	Prod. No. 0.00e+00;		
Matches 995;	Conservative	0;	Mismatches 5;	Indels 0;
			Gaps	0
CC	two alternative splicing mutations. These mutations establish the			
CC	cell phenotype of the ced-3 gene, confirming that ced-3, like ced-4,			
CC	function is not essential for viability. The ced-3 and ced-4 gene			
CC	products may be used to develop agents for treating conditions			
CC	characterised by cell deaths, such as myocardial infarction, stroke,			
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic			
CC	infection, aging or hair loss.			
SC	Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;			
Db	1801 ttccagctctgcaaacagaaacaaacacacacaaactttaaaatcagtttcaat	1860		
Qy	1801 TTCAGGCTGACAAACAAACAAACAAACAAACAACTTTTAAAAATCACTTTTCAAT	1860		
Db	1861 taaaaatacgaattctcatgtgaanaattgtgtttatgttttcgcaaaataaagaagaa	1920		
Qy	1861 TAAATAATACGATTCTCATTTGAATAATGTGTTTATGTTCGCAAAATTAAGAGAACT	1920		
Db	1921 gattcaaaaacaattttaacaaaaaaacccccaaatcgcgcagaataccaagatataaaa	1980		
Qy	1921 GATTTAAAAACAATTTTAAACAAAAAAAACCCCAAAATTCGCAGAAATCAAGATATAAAA	1980		
Db	1981 ttcaagagagtcataaatttcgcgattctacgtacttcaacttttttcgtagtcaagt	2040		
Qy	1981 TTCAAGAGAGTCATAATTTTCGAGATTTTACGACTTTACCTTTTTTTTCGTAGTTCACT	2040		
Db	2041 gcaagttgttgaggttttttcgcgaaaactaggaanaaaatcgataaaaattctcaaatcg	2100		
Qy	2041 GCAGTTGTGTGGAGTTTTCAGCAAAACACAGGAAAAATTCGATAAAAATTTACTCAAAACG	2100		
Db	2101 acctgaattttagagacaatgttttaaaaaaaacacbtltttccaatatttcaact	2160		
Qy	2101 AACTGATTTTGAAGACATGTTTAAAAAAAACCTATTTTTCCAATTAATTTTCACTCAT	2160		
Db	2161 ttccgactcaaatcgaaaatacaaatcgcgtactctgactacgcgctcagtagagagtcacac	2220		
Qy	2161 TTTCAGACTCAATTCGAAATCAAAATCGATCTGACTACGGGTCAGTAGAGAGTCACACC	2220		
Db	2221 atcagccgaagaatgtgcgtcaagatagaaggagcttgcctagagaggaacatlatagatg	2280		
Qy	2221 ATCAGCCGAGATGTGCTCTCAAGATAGAGAGAGCTTCTAGAGGGAACATTAATGATGT	2280		
Db	2281 tctctagtcataataaagtcgaatgtctctgaagttctctcgcgaacaacagtttga	2340		
Qy	2281 TCTCTAGTCATTTAAAAGTCGATGAATTTCTCGAAGTTCTCATGCCAAACAAGTCGTGA	2340		
Db	2341 atagtgataatagagataltgalttaatgtgagtttttaatcgataataaatttaaaaaa	2400		
Qy	2341 ATAGTGATTAATGAGATATGATTATATGTGAGTTTAAATCGAATTAATTTTAAAAAAA	2400		
Db	2401 aattgataataaagaatatatttttcgcgctatgttggaacggttcgcgagaagaagagg	2460		
Qy	2401 AATTGATTAATAAAGAAATATTTTTCGACTCTGTTGGAACGGTTTCGCGAAGAAACACGG	2460		
Db	2461 agatcgtgaagcaatgtgcaacgacgagagatgtgtgccttccacggttttlatgtgtc	2520		
Qy	2461 AGATCTGTAACACAGTCAGCAACGCGGAGAGATGTGGCTTCGACCGCTTTTATGATGCTC	2520		
Db	2521 ttgcgctctacggaacaggaagactgtctgaagttctctgaagctctgcgcagatcgttag	2580		
Qy	2521 TTGCGTCTCTCGGAGACAGAGAGACTTGTCTGAGAGTTCTTGAAACCTCTCCGCCAATGTAGG	2580		
Db	2581 tttttaaagttcggcgcaaaagcaagggtctcaacgaaaaaagagcgcgatcgttaattt	2640		
Qy	2581 TTTTAAATTTGGCGCAAAACCAAGGGTCTCACGAAAAAAGAGCGGATCGTAAATTTT	2640		
Db	2641 gcaaccacacggcaaggttttttccctccgaaaaatcggaaatattgcaactttcccaaat	2700		
Qy	2641 GCACCCACCGCACGGTTTTTCTCTCCGAAATGTGAAATTAATGACACTTTCGCAAAATAT	2700		

Query Match	99.48;	Score 994;	DB 9;	Length 7653;
Best Local Similarity	99.78;	Pred. No. 0.00e+00;		
Matches 997;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0

D	b	2581	ttttaagttcgcgcgaagaacgaaggcttcacgcaaaaaaagaagccgatcgtaattt	264
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat at 1490-1614"	
F	T		*tag= a	
F	T		/number= Exon_1	
F	T		2367..2429	
F	T		*tag= d	
F	T		/number= Intron_1	
F	T		2450..2575	
F	T		*tag= e	
F	T		/number= Exon_2	
F	T		2576..2853	
F	T		*tag= f	
F	T		/number= Intron_2	
F	T		2854..3107	
F	T		*tag= g	
F	T		/number= Exon_3	
F	T		3108..4302	
F	T		*tag= h	
F	T		/number= Intron_3	
F	T		3126..3243	
F	T		*tag= i	
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat at 3329-3396"	
F	T		3329..3396	
F	T		*tag= j	
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat at 3126-3243"	
F	T		3487..3759	
F	T		*tag= k	
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat at 3782-4070"	
F	T		3782..4070	
F	T		*tag= l	
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= m	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= n	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= o	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= p	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= q	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= r	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= s	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= t	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= u	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		4303..4634	
F	T		*tag= v	
F	T		/repeat_unit	
F	T		repeat_unit	
F	T		exon	
F	T			
F	T		/rpl_type= INVERTED	
F	T		/note= "Inverted w.r.t. repeat 3487-3759"	
F	T		43	

23-JUN-1994 (first entry)
 DE ced-3 gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT repeat_unit 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT 2310
 FT /tag= d
 FT /note= "C>T, from allele n1040"
 FT 2367..2429
 FT /tag= e
 FT /number= Intron_1
 FT 2430..2575
 FT /tag= f
 FT /number= Exon_2
 FT 2487
 FT /tag= g
 FT /note= "G>A, from allele n718"
 FT 2576..2853
 FT /tag= h
 FT /number= Intron_2
 FT 2854..3107
 FT /tag= i
 FT /number= Exon_3
 FT 3108..4302
 FT /tag= j
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= m
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= n
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= o
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= p
 FT /number= Intron_4
 FT 4688..4719
 FT /tag= q
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= r
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /tag= s
 FT /number= Exon_5

mutation 5757
 FT /tag= t
 FT /note= "G>A, from allele n2433"
 FT 5761..5814
 FT /tag= u
 FT /number= Intron_5
 FT 5815..5942
 FT /tag= v
 FT /number= Exon_6
 FT 5940
 FT /tag= w
 FT /note= "C>T, from allele n1165"
 FT 5943..6297
 FT /tag= x
 FT /number= Intron_6
 FT 6062..6138
 FT /tag= y
 FT /rpt_type= INVERTED
 FT 6297
 FT /tag= z
 FT /note= "G>A, from allele n717"
 FT 6298..6537
 FT /tag= aa
 FT /number= Exon_7
 FT 6322
 FT /tag= ab
 FT /note= "C>T, from allele n1949"
 FT 6372
 FT /tag= ac
 FT /note= "G>A, from allele n1286"
 FT 6434
 FT /tag= ad
 FT /note= "C>T, from alleles n1129 and n1164"
 FT 6485
 FT /tag= ae
 FT /note= "C>T, from allele n2430"
 FT 6535
 FT /tag= af
 FT /note= "G>A, from allele n2426"
 FT 6538..7012
 FT /tag= ag
 FT /number= Intron_7
 FT 6567..6625
 FT /tag= ah
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /tag= ai
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7013..7075
 FT /tag= aj
 FT /number= Exon_8
 FT 7020
 FT /tag= ak
 FT /note= "C>T, from allele n1163"
 FT mutation
 PN M09325685-A.
 PD 23-DEC-1993. U05701.
 PE 14-JUN-1993; US-897788.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shiham S, Yuan J;
 DR WPI, 94-007542/01.
 DR P-PSDB; R47466.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 2; Fig 4; 127pp; English.
 CC This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
 CC was identified as the ced-3 transcript and was most abundant in
 CC embryos, but was also detected in larvae and young adults, suggesting
 CC that ced-3 is expressed not only in cells undergoing cell death. The
 CC four largest introns as well as sequences 5' of the start codon

D	1801	ttccagcctgacaaacgaacaacaaaccacaacatltttaaataatccagtttccaat	1860
Q	1801	ttccagcctgacaaacgaacaacaaaccacaacatltttaaataatccagtttccaat	1860
D	1861	taaaaaaacagatcttccattgaaaatctgcttttttggttgcgaataataaagagact	1920
Q	1861	TAAAAATAAACGATTTCTCATTTGAAAATTTGTGTTTATGTTTTCGAAAAATAAAGGAAC	1920
D	1921	gattcaaacacatltttaaacaataaaaccccaaaatltccgcagaaatcaagatlaaaaaa	1980
Q	1921	GATTCAAAACATTTTAAACAAAAAAACCACAAATTCGCCAGAAATCGAAGATTAATAA	1980
D	1981	ttcaagaggctcaaaatlttcgaatlttaccgtacttcaaccttlttctgtagtcaat	2040
Q	1981	TTCAAGAGGGTCAAAATTTTCCCATTTTACTGACTTCACCTTTTTCGTAAGTTCAGT	2040
D	2041	gcattcttggaggttttttgcgaataactggaataaaatcgaataaaatctactcaatcg	2100
Q	2041	GCACTTGTGGAGTTTTTGCGAAACTGGAATAAATCGATTAATAATTTACTCAAAATCG	2100
D	2101	agcgaattttgagagacaagtlttaaaaaaacactatlttccaataatctcaat	2160
Q	2101	AGCGATTTTGGAGCAATGTTTAAAAAAAACATATTTTCCATATATTTCATCTCAT	2160
D	2161	ttcagaactaatcgaaaatcacaactcgactctgactacggttcagtagaaggtcaac	2220
Q	2161	TTTCAGACTTAATCGAAAATCAATATGTACTCGACTGACTACGGGTCAGTAGAGGTCAAC	2220
D	2221	atcagccggaagatgagtcgccaagatagaagaagcttctgaagaaggaaatcatagat	2280
Q	2221	ATCCGCCGAAGATGATGCGTCAAGATTAAGAAAGGAGCTGTCTAABAGAGGAACATTAATGATG	2280
D	2281	tctctagtcactaaagctgcagtcaactctcgaagtctcgaatctcgaataaacaagttga	2340
Q	2281	TCTTACTCATCTTAAGTGCATGAAATTCGCAGATTCCTCATTCGCAAAACAAGTGTGA	2340
D	2341	atagtgataatgagatatgatcaatgtagtattttaatcgataataatlttaaaaaa	2400
Q	2341	ATAGTGATTAATGAGATATGATTAATGATGAGTTTAAATCGAATATATATTTAAAAAA	2400
D	2401	aattgataataataagaataatlttttcagtcatactgtagaacggttcgcgagaagagacgg	2460
Q	2401	AATTGATTAATATAAAGATATTTTTCAGTCATGTCGAACGGTTCGCGAAGAGAACGGG	2460
D	2461	agatcgtgaaagcagtcgcaacgcggaagagatgtgacgttcgaacggtttatgtagtgc	2520
Q	2461	AGATCGTGAAGCACTGCCAACGACGGGAGATGTGCGCTTCGACGCGTTTATGATGCTC	2520
D	2521	ttcgcttaccggagacgaagaagacttgcgtgaagtlttctaacactctcgcagaatcgttag	2580
Q	2521	TTCGCTTACGGGACACAGAAAGACTTGTGTAAGTCTTAACCTTCGCGCAATGCTAGG	2580
D	2581	tttttaaatgttcgagcaaaacgaaggttctcagcaaaaaaagagcggaatcgttaattt	2640
Q	2581	TTTTTAAGTTCGGCGCAAAAGCAAGGAGTCTACGCGAAAAAAGAGCGGATGCTAATTTT	2640
D	2641	gcaacccacgcgacggttttttcctccgaataatcgaaatatgacacttcccaaat	2700
Q	2641	GCAACCCACGCGCGGTTTTTTCCTCCGAATAATCGGAATTAATGACTTTTCCCAAAAT	2700
D	2701	ttgaagtgaataatatttatttaccgaaagctcgaagtatattattttaaacta	2760
Q	2701	TTGAAGTGAATATATTTTATTACGAAAGCTCGAGTGATTTATTTTAAACACA	2760
D	2761	atttcgtgagcgaagaagccatttgttagatttgcgaa	2800
Q	2761	ATTTCGTGAGCGCAAAAGCGCATTTGTGATTTGCGCAA	2800
RESULT 3			
ID 054666 standard: DNA: 7653 BP.			
AC 054666;			

```

|||||
QY 1801 TTCCAGGCTGCAACAGAAACAAACACACACAAATTTTAAATTCAGTTTCAAT 1860
Db 1861 taaaaataacgatttccatctgtaaatgtgtttatgttgcgaataataaagagact 1920
QY 1861 TAAAAATACGATTTCATGTGAATAATGTGTATTAGTTGGCAAAATTAAGAGAACT 1920
Db 1921 gatccaaacacattttaacaaaaaaacccccaaatcgcggaataatcaagataaaaaa 1980
QY 1921 GATTCAAAACAAATTTTAAACAAAAAAACCCCAAAATTCGCAGAAATCAAGATAAAAA 1980
Db 1981 ttcaagaggggtcaaaatttccgaatttctactgacttccacatttcttcgttagtca 2040
QY 1981 TTCAAGAGGGTCAAAATTTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCA 2040
Db 2041 gcaagtgttggagtttttgcgaaaaacttaggaaaaaaatcgataaaatactcaatc 2100
QY 2041 GCAGTGTGTGGAGTTTGTGACGAAACACTAGGAAAAAATCGATAAATTAATCAATCG 2100
Db 2101 agcgtgaatttggagacaatgtttaaaaaaaacacataatttccataatttcactcat 2160
QY 2101 AGCTGAATTTTGGAGACAATGTTTAAAAAAACACATAATTTTCCAAATTAATTCAC 2160
Db 2161 ttccagactaaatcgaaaaatcaaatcgttactctgactcagcggtcagtagagagtc 2220
QY 2161 TTTGAGACTAAATCGAAATTAATTCGTACTCTACTACGAGGTAGTAGAGAGTCAACC 2220
Db 2221 atcagccgaagatgatgcgtlcaagaatagaagaagctgtctagagaaggaataatgat 2280
QY 2221 ATCAGCCGAAGATGATGGGTCAAGATAGAGAGAGCTTGCATAGAGAGACATATATGAT 2280
Db 2281 tctctagtcataaagtgcgaatgtcgaatctcgaattctcactgcgaacaaagtgltga 2340
QY 2281 TCTCTAGTCACTAAAGTGCATGAATTCGAAATTCGAAATTCATCGCAAAACAAGTGTGA 2340
Db 2341 atagtgataatgagatgatgttaattgagttttaaatacgaataaataattttaaataa 2400
QY 2341 ATAGTGAATATGAGATATGATATGATGATTTTAAATCAATTAATTAATTAATAA 2400
Db 2401 aatgataataataagaatatttgcagtcagtgtgaaagctgtcgcggaagaagacgag 2460
QY 2401 AATGATATATTAAGAATATTTTTCAGTCATGTGAGACGGTTCGCGAGAGAGACGGG 2460
Db 2461 agatcgtgaaagcagtcgaacgcgggagagatgtgcgcttcgcagcgttttaagatgctc 2520
QY 2461 AGATCGTGAAGACAGTCAACGAGCGGAGATGTGGCGTTGAGCGCTTTATGATGCTC 2520
Db 2521 ttgcgtctacgcgcacgaaagacttgcgaagtcttgaacctctgcgcagatcgtagg 2580
QY 2521 TTCGCTCTACGCGACACGAGAGACTTGTGAGTTCTTGAACCTCTGCGCAGATCGTAG 2580
Db 2581 ttcttaaatctcgcgcgaacgaaggtctcgcgaaaaaaagagcggaatcgttaatttc 2640
QY 2581 TTTTAAAGTTCGCGGCAAAAGCAAGGGTCTCAGGAAAAAAGGCGGATCGTAATTTT 2640
Db 2641 gcaaccacacgcggcgttttctccgcgaataatcgaaatattgcacttcccaatat 2700
QY 2641 GCAACCCACCGCGAGGTTTTCCTCCGAAATCGGAAATTAATGACATTTCCCAATAT 2700
Db 2701 ttgaagtgaaatatattttactgaagctcgaatgattatttttttaacacta 2760
QY 2701 TTGAAGTGAATATATTTTATTTACTGAAAGCTCGAATGATATTTTATTTTAACACTA 2760
Db 2761 atttcgtgcgcgaagaagccattttagattgcgaa 2800
QY 2761 ATTTTCGTGGCGCAAAAGCCATTTTGTAGATTGCCGAA 2800

```

RESULT 2
 ID T38196 standard: DNA; 7653 BP.
 AC T38196;
 DF 17-Dec-1996 (first entry)
 DE Nematode Ced-3 gene.

```

KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;  

KW cell death; apoptosis; neural degeneration; inflammation;  

KW antiinflammatory; ds.  

FH Caenorhabditis elegans.  

FH Key Location/Qualifiers  

FT repeat_unit 1356..1472  

FT FT  

FT /tag= a  

FT /label= Repeat-1  

FT 1490..1614  

FT /tag= b  

FT /label= Repeat-2  

FT 2167..2366  

FT /tag= c  

FT /codon_start= 2232..2234  

FT 2310  

FT /tag= d  

FT /label= T(n1040)  

FT /note= "causes L27F mutation"  

FT 2367..2429  

FT /tag= e  

FT /label= Intron-1  

FT 2430..2575  

FT /tag= f  

FT 2487  

FT /tag= g  

FT /label= A(n718)  

FT /note= "causes G65R mutation"  

FT 2576..2853  

FT /tag= h  

FT /label= Intron-2  

FT 2854..3107  

FT /tag= i  

FT 3108..4302  

FT /tag= j  

FT /label= Intron-3  

FT 3126..3243  

FT /tag= k  

FT /label= Repeat-1  

FT 3329..3396  

FT /tag= l  

FT /label= Repeat-1  

FT 3487..3759  

FT /tag= m  

FT /label= Repeat-2  

FT 3782..4070  

FT /tag= n  

FT /label= Repeat-2  

FT 4303..4633  

FT /tag= o  

FT 4634..5546  

FT /tag= p  

FT /label= Intron-4  

FT 4688..4719  

FT /tag= q  

FT /label= Repeat-3  

FT 5221..5330  

FT /tag= r  

FT /label= Repeat-3  

FT 5546..5760  

FT /tag= s  

FT 5737  

FT /tag= t  

FT /label= A(n2433)  

FT /note= "causes G360S mutation"  

FT 5761..5814  

FT /tag= u  

FT /label= Intron-5  

FT 5815..5942  

FT /tag= v  

FT 5940  

FT /tag= w  

FT /label= T(n1165)  

FT /note= "creates premature stop codon at 403"

```

```

FT      /*tag= g
FT      /note= "G>A, fromm allele n178"
FT      2576..2853
FT      /tag= h
FT      /number= Intron_2
FT      2854..3107
FT      /tag= i
FT      /number= Exon_3
FT      3108..4302
FT      /tag= j
FT      /number= Intron_3
FT      3126..3243
FT      /tag= k
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT      3329..3396
FT      /tag= l
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT      3487..3759
FT      /tag= m
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT      3782..4070
FT      /tag= n
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT      4303..4634
FT      /tag= o
FT      /number= Exon_4
FT      4635..5546
FT      /tag= p
FT      /number= Intron_4
FT      4688..4719
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5221..5330
FT      /tag= r
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5547..5760
FT      /tag= s
FT      /number= Exon_5
FT      5757
FT      /tag= t
FT      /note= "G>A, from allele n2433"
FT      5761..5814
FT      /tag= u
FT      /number= Intron_5
FT      5815..5942
FT      /tag= v
FT      /number= Exon_6
FT      5940
FT      /tag= w
FT      /note= "C>T, from allele n1165"
FT      5943..6297
FT      /tag= x
FT      /number= Intron_6
FT      6062..6138
FT      /tag= y
FT      /rpt_type= INVERTED
FT      6297
FT      /tag= z
FT      /note= "G>A, fromm allele n717"
FT      6298..6537
FT      /tag= aa
FT      /number= Exon_7
FT      6322
FT      /tag= ab
FT      /note= "C>T, from allele n1949"
FT      6372
FT      /tag= ac

```

```

FT      /note= "G>A, from allele n1286"
FT      6434
FT      /tag= ad
FT      /note= "C>T, fromm alleles n1129 and n1164"
FT      6485
FT      /tag= ae
FT      /note= "C>T, from allele n2430"
FT      6535
FT      /tag= af
FT      /note= "G>A, from allele n2426"
FT      6538..7012
FT      /tag= ag
FT      /number= Intron_7
FT      6567..6625
FT      /tag= ah
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      /tag= ai
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7013..7075
FT      /tag= aj
FT      /number= Exon_8
FT      7020
FT      /tag= ak
FT      /note= "C>T, from allele n1163"
FT      1..5850
FT      /tag= al
FT      /note= "Claim 2, inhibitor fragment"
FT      1..3020
FT      /tag= am
FT      /note= "Claim 2, inhibiotr fragment"
FT      m90325694-A.
FT      23-DEC-1993.
FT      14-JUN-1993: US-897788.
FT      12-JUN-1992: US-984182.
FT      20-NOV-1992: US-984182.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Horvitz HR, Shaham S, Yuan J;
FT      WPI; 94-007551/01.
FT      P-PSDB: R45262.
FT      Agents which affect activity of cell death genes - used to
FT      develop drugs for treating conditions characterised by cell death
FT      or proliferation
FT      PS Disclosure: Fig 3: 132pp; English.
FT      CC This sequence encodes the C.elegans cell death gene, ced-3. Fragments
FT      of the amino terminal of the protein encoded by this sequence act as
FT      inhibitors of ced-3. This gene has considerable similarity to human
FT      interleukin-1beta convertase (ICE), which converts pro-interleukin-
FT      1beta to the active cytokine and is involved in inflammatory response
FT      in humans. The similarity between the two sequences suggests that
FT      inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
FT      nematode Ced-3 proteins have an overall amino acid similarity of 28%.
FT      CC The ced-3 inhibitors may be used for identifying agents which affect
FT      the activity of a gene belonging to the to the ced-3/ICE family of
FT      genes and for diagnosis of diseases characterised by cell death. They
FT      can also be used to develop drugs for treating conditions characterised
FT      by cell deaths such as myocardial infarction, stroke, degenerative
FT      disease, traumatic brain injury, hypoxia, pathogenic infection, or
FT      hair loss, or drugs for reducing the proliferative capacity or size
FT      of a population of cells such as cancerous cells, cells which produce
FT      autoreactive antibodies, infected cells, hair follicle cells or cells
FT      which are critical to the life of a parasite, pest or recombinant
FT      organism. They may also be used in the diagnosis of inflammatory
FT      disease.
FT      CC Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
SQ
Query Match 100.0%; Score 1000; DB 9; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1801 ttccagctgcataacagacaacacacacattttaaaatcagtttccaat 1860

```

RESULT	1	
ID	054401	standard: DNA: 7653 BP.
AC	054401;	
DT	01-JUL-1994	(first entry)
DE	ced-3 gene.	
RW	C.elegans; cell death; gene: ced-3; inhibition: human: parasite;	
RW	interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;	
RW	inflammatory response; nematode; diagnosis; myocardial infarction;	
RW	stroke; degenerative disease; traumatic brain injury; hypoxia; pest	
RW	pathogenic infection; hair loss; cancer; autoreactive antibody; ss.	
OS	Caenorhabditis elegans.	
FT	Key	Location/Qualifiers
FT	Repeat_unit	1356..1472
FT		/*tag= a
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 1490-1614"
FT	Repeat_unit	1490..1614
FT		/*tag= b
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 1356-1472"
FT	exon	2232..2366
FT		/*tag= c
FT		/number= Exon_1
FT	mutation	2310
FT		/*tag= d
FT		/note= "C>T, from allele n1040"
FT	intron	2367..2429
FT		/*tag= e
FT		/number= Intron_1
FT	exon	2430..2575
FT		/*tag= f
FT		/number= Exon_2
FT	mutation	2487

Query Match	4.2%;	Score 42;	DB 19;	Length 145920;
Best Local Similarity	75.6%;	Pred. No. 5.20e-06;		
Matches	62;	Conservative	0;	Mismatches 20;
			Indels	0;
			Gaps	0

```

Db 93157 GCACTTCTCCAAATATTTCAGTGGAAAGCATTTTATTTTCTGGTAAATCAAGGATTTT 93216
      ||||| | ||||| | ||||| ||| | | ||||| |
QY 2685 GCACTTCCCAAATATTTGAAGTGAATATATTTTATTCTGAAAGCTCAGAGATTTAT 2744

```

Db	93217	TTATATTTTGAGACTATTTTC	93238
QY	2745	TTATTTTTTAACACTAATTTTC	2766

RESULT	15			
LOCUS	CEY56A3	224746 bp	DNA	HTG 30-DEC-1998
DEFINITION	Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from clone Y56A3, WORKING DRAFT SEQUENCE.			

NID	g4090202
VERSION	AL022280.1 GI:4090202
KEYWORDS	HTG: HTGS_PHASE1.
SOURCE	Caenorhabditis elegans
ORGANISM	Caenorhabditis elegans

ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditioidea; Rhabditidae; Pelodermiae; Caenorhabditis
REFERENCE
1 (bases 1 to 224746)

TITLE Direct Submission
ORIGINAL Submitted (30-DEC-1998)
PROJECT Nematode Sequencing Project, Sanger Centre

COMMENT
O Dec 31, 1998 this sequence version replaced g1:4006996.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

- * NOTE: This is a 'working draft' sequence.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

FEATURES	Location/Qualifiers
source	1. .224746

[illegible]

Query Match	4.28;	Score 42;	DB 19;	Length 224746;
Best Local Similarity	92.08;	Pred. No. 5.20e-06;		
Matches	46;	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0

DB 105007 AAAAAGACGCGAGCGTAATTTCCGAACCCTGGCGCACGGTTTTTTCCTC 105056
|||||
|||
QY 2618 AAAAAGACGCGATCGTAATTTTGCAACCCACCGGCACGGTTTTTTCCTC 2667

Search completed: Fri Aug 6 03:41:45 1999
Job time : 1847 secs.

C. elegans cDNA yk170a12.5"

```

/codon_start=1
/protein_id="AAB42290.1"
/db_xref="GI:1825673"
/translation="MSRFLSCLLAALVATSATSAOFTDOCTODYAAITTCQYVYVMGY
NISMTPVPEVYSFHRARDLIDADGLIOPVYCOLGNALNSCLAPYSCGPNAYMOM
MAANTTEADYWDLAATVEYOCGVYNTLMTPEYCMACFCRDYOPNIDOCDAQALDM
KMGMPCAAOQDFNCOAAVYRNCCDNMAGVYICNVYASGKAVNPACVNAAGLVTCDA
PR"
complement(3730..6640)
/gene="F53E10.5"
complement(join(3730..3855,3955..4242,4314..4403,
4468..4728,5527..5694,5875..5931,6106..6195,6269..6332,
6435..6518,6572..6640))
/gene="F53E10.5"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42291.1"
/db_xref="PID:g1825674"
/db_xref="GI:1825674"
/translation="MDNNSNVEPTEPELEDELFIKTNMTNEMYOHAIVYKTNNEADAE
TLNVEKMDPHDNDNDNDPEIROPPEIDRPPSYTVNENHDDPYOBELETTATEE
NTPSSREPSTNGHYNDGNAROSTPOFNDAEIAPIISVPLIAHNSQIHFENTLP
VFOHVMHICQANQPEKPRHRIKTELMDLIGTLQOKARINTEITENKENTIE
LOKSTTAERLNDLKNNAKPAENDEKMTIKGTGSRKTRAKQSESTTLAYKIRLGE
VETIKQALSETLEOIAKNAVOAENGIPANVYERPLHTEDRSDVSHALDPTCSLL
PARRHITGSNRAQONMAIASIPO"
complement(9953..10770)
/gene="F53E10.6"
complement(join(9953..10211,10265..10413,10478..10566,
10692..10770))
/gene="F53E10.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42292.1"
/db_xref="PID:g1825675"
/db_xref="GI:1825675"
/translation="MSTKNRDLRVLTEDSDDDNEREMSSGGEGEGPSSVDGAGD
ADETVAPATERKKVKIKTKKQSLKSVKYEIKALVAPDITTDREKEMIRR
VATGCVOLFNVAUSDROKTMADAKEMTAREAREAROFEDKNFSDRFDASGVYGA
KKEYKGEDDDEGDMDIGEPOIDTGNSSDD"
12240..12611
/gene="F53E10.3"
join(12240..12318,12363..12562,12698..12767,15605..15662,
15709..15811)
/gene="F53E10.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42289.1"
/db_xref="PID:g1825672"
/db_xref="GI:1825672"
/translation="MAETKMPSSKNSISISIEGSLNSVSSHVVKPEIIGIRPRFLKT
KLIFRASANTKNNDPEKMSALYSFCSLHRLDFSGDFVEMTHLKKFFPK
FVFDAPYELFHFPHQHLFRNFQKIIFKNLFRKRLSEALNNEFPVYKFLKLSIF
FYEGMSTYS"
16384..21870
/gene="F53E10.2"
join(16384..16434,16487..16554,17235..17462,17579..17656,
17699..17860,18899..19194,19511..19653,20539..20673,
21691..21870)
/gene="F53E10.2"
/note="similar to acetylcholine receptor protein"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42293.1"
/db_xref="PID:g1825676"
/db_xref="GI:1825676"
/translation="MISQQLIRIVLHFLMGLAMAOOKLIKLYFRDYKRELSPYKD
ESSGPTNIVVQYFOYKQKVENDDIITXCMIEYMODEFLTNPDGNSISLHAP
SEMIMKPDLLVYNANMNKEMQNTVOIEHTGKISFRALITDIDCDLOMEKPPD
OQICFIMLASWSGSOIMLNTAEQPTAEPALDKITLALNITRNEMKLYDFEY
RNLLKYDDCCPNPDPDISYFAIKRNPSTYLTLLIPSAFTIVVIGFPPHSSTGE

```

NTKVSIGYVALLSLAIIIMMSDKLPATNSVPLIGOYYIGLIFIMPLATCTETEL
GIOMGNAGRPITRRLRSFLSLTRNNNSVQVFPFERELMNDQESIKMRLLKRYHNKTA
SLIODESIDSKLDPAMPYVNVVSVISVQISKODLLSOHLKRIKEMOMLAR"

gene
23195..26175
/gene="F53E10.1"
join(23195..23587,23688..23798,23847..23935,24914..25060,
25114..25297,25708..25788,25837..26175)
/gene="F53E10.1"
/note="coded for by C. elegans cDNA yk167d12.3; coded for
by C. elegans cDNA yk167d12.3; coded for by C. elegans
cDNA yk60a3.5; coded for by C. elegans cDNA yk169e10.5"

CDS
/codon_start=1
/protein_id="AAB42288.1"
/db_xref="PID:g1825671"
/db_xref="GI:1825671"
/translation="MLELDKDKVYLICLASSFIYHTRAQIYSYSAKREWPYGI
VGLHIYPIKSCKPDVAFKCTKLGMELEDRVVLVEESTGKTRIRQPKLVN
VESYMDGILEISVGHKLSVDLKRVENGRITRAITFNLQODGDEYAKLLS
DYIEEPNRYLLFSKGLYTERTCIPDOQNTTPYKRRDNGSFTDLAPFLIATEASLK
AVNEKLDKVMRMRNRPSTIYEGCADEDEKMAEIRIGEAHLECFACRCLVTVDP
EYGESEKEMQPLKTRERLAPPEGKMSAKHDSVPFGVYAGTVNAEVIHIGTAAYVYK
KPTVE"

BASE COUNT 7844 a 4906 c 4801 g 8725 t

ORIGIN

Query Match 4.2% Score 42; DB 21; Length 26276;
Best Local Similarity 92.08; Pred. No. 5 20e-06;
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 20234 AAAGAGCGGAGCGTAAATTTGCAACCTGGCGGACGGTTTCTTCC 20283
Oy 2619 AAAGAGCGGAGCGTAAATTTGCAACCGGCGGCGGTTTCTTCC 2668

RESULT 14
LOCUS AC006710 145920 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone y119C1AA, WORKING DRAFT SEQUENCE, 1
unordered pieces.
AC006710
AC006710
AC006710.1 GI:4263228
VERSION HTG; HTGS_PHASE1.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 145920)
Waterston, R.H.
The sequence of Caenorhabditis elegans clone
unpublished
2 (bases 1 to 145920)
Waterston, R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 145920: contig of 145920 bp in length.
Location/Qualifiers
1..145920
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y119C1AA"
BASE COUNT 46442 a 26006 c 25762 g 47710 t
ORIGIN

FEATURES
source

The true left end of clone F35C5 is at 1 in this sequence. The true right end of clone F35C5 is at 8095 in sequence 281556.
The true left end of clone F58G1 is at 33064 in this sequence. The start of this sequence overlaps with the end of sequence 292852.
The end of this sequence (33064..33164) overlaps with the start of sequence 281556.

Location/Qualifiers

```

1..33164
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="F35C5"
2293..3419
/gene="F35C5.1"
/join(2293..2396,2761..2934,3207..3326,3374..3419)
/gene="F35C5.1"
/codon_start=1
/protein_id="CAB03050.1"
/db_xref="PID:e1346293"
/db_xref="PID:g3876676"
/db_xref="GI:3876676"
/db_xref="SPTREMBL:O45437"
/translation="MCHLHLSMTFLFPLILVLYLVYVCSKSKOKRIDSLKVPBR
PPOASMSGSKSYRSGKNGOVSVRGSDTDSAKETLEMPGRKARADPEKRNILIT
PAAROKTGDSSSPSKRYOFAPHLKRYKRAITMPEDADEKMTITA"
6734..10069
/gene="F35C5.2"
/join(6754..6993,7040..7185,7766..7841,7890..8089,
8366..8552,9216..9256,9358..9410,9879..10069)
/gene="F35C5.2"
/codon_start=1
/protein_id="CAB03051.1"
/db_xref="PID:e1346294"
/db_xref="PID:g3876677"
/db_xref="GI:3876677"
/translation="MSEECPEPKASDDOILLOTSMILRLNITCTFLSIGCPYAVAC
IRPMKHPHIFESSTALILYLAFVDFVAVHVSOMAVYRSIFYADDCNITFSD
DCLSLGRIVFGISGLIFIRHSLSDGLAFTTFYFKQORMGCIILAYLMTSVSG
VOFLPLIGESKEDLPSCQFRKQDAPRANLFTISSLVLTINLWVNLFLFVKNKH
SKRTFDEVFQYORSEAMTSEISLVIAQISAGIYAGSMLFROAKBEIPYLLS
NLIIWVAISGFIIEHFKLFGIAHPYNNKNSGTAGNLIRYRNPPTSHHILHSIC
KTAPADHSHHKPGQEPGSONAGAAALGIIVLWFLCEILK"
10256..11938
/gene="F35C5.3"
/join(10256..10321,10369..10455,10501..10557,10605..11249,
11302..11355,11411..11455,11506..11547,11595..11630,
11678..11710,11762..11827,11885..11938)
/gene="F35C5.3"
/notes="cDNA EST yk207f10.3 comes from this gene: cDNA EST
yk207f10.5 comes from this gene"
/codon_start=1
/protein_id="CAB03052.1"
/db_xref="PID:e1346295"
/db_xref="PID:g3876678"
/db_xref="GI:3876678"
/db_xref="SPTREMBL:O45438"
/translation="WKSLVFLFVAFKKSADTAPVDFNKIEVTEVELMDAINQLIL
EOMGAIRSTTPIQSIPIITADPTADTAETPAETKAEITVAQEAETSEAEHETIV
GQSDTVRSGESQETTSQITVSOENTVSOETTSQITVSOETTSQITVSOETTSQITV
SOETTSQITVSOETTSQITVSOETTSQITVSOETTSQITVSOETTSQITVSOETTSQITV
TSOETTSQITVSOETTSQITVSOETTSQITVSOETTSQITVSOETTSQITVSOETTSQITV
TTAVYRGSDDPCDTTEATFGSTYGESPRCPTTASSEPRTGHTDOAEVATVPASQATIS
ATIIIGVDGTGSLNLEATTEQICKHPTKDSSEKTEDEKDKETTRAGTVNHRPERDI
LK"
12108..13502
/gene="F35C5.4"
/join(12108..12128,12366..12551,12916..12993,13048..13095,
13141..13185,13241..13279,13338..13379,13443..13502)
/gene="F35C5.4"
/codon_start=1
/protein_id="CAB03053.1"
/db_xref="PID:e1346296"

```

```

/db_xref="PID:g3876679"
/db_xref="GI:3876679"
/translation="MGASTAIVAIIOSEPSTSSSSSSSPSSRPEDGATVGNFTOSSAV
SSQSTSESRSSVSPDSPVPTSTLLSYGQVNVNLPFISQNPANVSBEKNPLSPIF
APTTTINLAPITVYPIPLTTITSLPIEQVYVNLPIFISQNPANVSBEKNPLSPIF
NDVPSAYOPLIN"
complement(14630..16576)
/gene="F35C5.5a"
complement(14630..16576)
/gene="F35C5.5b"
complement(join(14630..14809,14862..14974,15032..15229,
15289..15584,15632..15735,16263..16376,16418..16576))
/gene="F35C5.5a"
/notes="similar to Lectin C-type domain short and long
forms, von Willebrand factor type A domain: cDNA EST
yk204f1.3 comes from this gene"
/codon_start=1
/protein_id="CAB03054.1"
/db_xref="PID:e1346297"
/db_xref="PID:g3876680"
/db_xref="GI:3876680"
/db_xref="SPTREMBL:O45440"
/translation="MIGILLIFLSVYGYPPTRNCGHDLANIMLVVYVYDNSA
PMTOBELTVAQIYTVBAGRTITNIDKTTTGYGLTYTNTATIODLNRRQSPD
DLSTVQILPNDLSTSEDFLAKIGAEOLLAGRRNNRNKRNKQWIVYASVND
EGEEDPRPFAERLAKGSVIAVAFDQDEEMILIEIATPGFNEDENAIKEI
OTAMIOITNCYCSNLWTOYRETFEDHSKFGVCLPVALTAGWPAKACOMIOTGY
MYTEDOKKHDFVFLVRNDTSFEPYVYHGLSYNGGYAMQOVGHALPNDLSK
NGFPPOSTTAVLNOOSSAFRGWNNVNYTYSERVYCAACDSTYCE"
complement(join(14630..14815,14862..14974,15032..15229,
15289..15584,15632..15735,16263..16376,16418..16576))
/gene="F35C5.5a"
/notes="similar to Lectin C-type domain short and long
forms, von Willebrand factor type A domain: cDNA EST
yk367h4.3 comes from this gene: cDNA EST yk367h4.5 comes
from this gene: cDNA EST yk363d9.3 comes from this gene:
cDNA EST yk363d9.5 comes from this gene: cDNA EST
yk204f1.5 comes from this gene"
/codon_start=1
/protein_id="CAB03055.1"
/db_xref="PID:e1346298"
/db_xref="PID:g3876681"
/db_xref="GI:3876681"
/db_xref="SPTREMBL:O45441"
/translation="MIGILLIFLSVYGYPPTRNCGHDLANIMLVVYVYDNSA
PMTOGLIEVAQIYTVBAGRTITNIDKTTTGYGLTYTNTATIODLNRRQSPD
DLSTVQILPNDLSTSEDFLAKIGAEOLLAGRRNNRNKRNKQWIVYASVND
EGEEDPRPFAERLAKGSVIAVAFDQDEEMILIEIATPGFNEDENAIKEI
OTAMIOITNCYCSNLWTOYRETFEDHSKFGVCLPVALTAGWPAKACOMIOTGY
MYTEDOKKHDFVFLVRNDTSFEPYVYHGLSYNGGYAMQOVGHALPNDLSK
LWNPGEPTOSTTAVLNOOSSAFRGWNNVNYTYSERVYCAACDSTYCE"
17118..18865
/gene="F35C5.6"
/join(17118..17665,17998..18194,18243..18553,18686..18865)
/gene="F35C5.6"
/notes="similar to Lectin C-type domain short and long
forms, von Willebrand factor type A domain: cDNA EST
CMS887f comes from this gene: cDNA EST yk279b9.5 comes
from this gene: cDNA EST EMBL:D69169 comes from this gene:
cDNA EST EMBL:D69482 comes from this gene: cDNA EST
EMBL:D69566 comes from this gene: cDNA EST EMBL:D69586
comes from this gene: cDNA EST EMBL:D69649 comes from this
gene: cDNA EST EMBL:D66470 comes from this gene: cDNA EST
EMBL:D70265 comes from this gene: cDNA EST EMBL:C11927
comes from this gene: cDNA EST EMBL:C10473 comes from this
gene: cDNA EST EMBL:C12088 comes from this gene: cDNA EST
EMBL:C11616 comes from this gene: cDNA EST EMBL:C13881
comes from this gene: cDNA EST yk207c12.3 comes from this
gene: cDNA EST yk207c12.5 comes from this gene: cDNA EST
yk234d2.3 comes from this gene: cDNA EST yk235f6.3 comes
from this gene: cDNA EST yk235f6.5 comes from this gene:
cDNA EST yk261g11.3 comes from this gene: cDNA EST

```

MEDLINE 97132596
 REFERENCE 2 (bases 1 to 10772)
 AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
 TITLE Structure of the cysteine proteinase (Cp1) gene of *Drosophila melanogaster* and associated mutational effects
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 10772)
 AUTHORS Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia

FEATURES
 source 1..10772
 location/Qualifiers
 1..10772 Drosophila melanogaster
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 join(872..1000,2310..2426,6476..6690,6751..7707)
 /gene="Cp1"
 /product="cysteine protease"
 872..7707
 /gene="Cp1"
 872..1000
 /gene="Cp1"
 /number=1
 1001..2309
 /gene="Cp1"
 2310..2426
 /gene="Cp1"
 /number=2
 join(2328..2426,6476..6690,6751..7462)
 /gene="Cp1"
 /codon_start=1
 /product="cysteine proteinase-1"
 /protein_id="AAB65749.1"
 /db_xref="GI:2305221"
 /db_xref="PID:92305221"
 /translation="MRTAVLLPLALLAAVAVSFADVYVMEEMTEKLEBRKYNODET
 EERFKIKENENKHKIAKNORAEKGVSEFKLVNKAADLLHEEFOLANNGFNLYLK
 QLEAASESKGVTFTSPAVHTLPKSYDMTKKAVTRVKQGGCGSMARSSGALGEO
 HFRKSGVLSLEQNLVDCSTKYGNNCGGLMDNFRITKONGIDTEKSTPIEYAD
 DSCHEFKGTGATDGRFTDIPQDEKKAEEAVATGVSVAIDASHESQFSEGVYN
 EPQCDANLDHGVVFGTDESDEGYLWVKNSWGTWGDGKFIKMLRNKENCGLIAS
 ASSPLY"
 2427..6475
 /gene="Cp1"
 4546..4553
 /gene="Cp1"
 /note="Insertion site of P[Casper](50C)"
 /citation=[1]
 6476..6690
 /gene="Cp1"
 /number=3
 6691..6750
 /gene="Cp1"
 6751..7707
 /gene="Cp1"
 /number=4
 join(8110..9300,9370..9532)
 /product="phenylalanyl tRNA synthetase"
 join(8110..9300,9370..9532)
 /note="potential orf"
 /codon_start=1
 /product="phenylalanyl tRNA synthetase"
 /protein_id="AAB65750.1"
 /db_xref="PID:92305222"
 /db_xref="GI:2305222"
 /translation="MLTLRVGARHMKSTRCLASSAPAKSPSSPOLLEVSGSTVA
 TDCWTVTEKILSYGANKRLQDHPLESLIRQIVYFYGATRNQNGNLFSEVYQDN
 PVTVQONEDNLIIPADHVSROKSDCYINQOHLRAHTTAHVEISGLDNFLVVG
 EYVRDEIDSTHYPVHADAVLVTKDLFERNPGLLEFEETSGTLADPKLILPH
 SSWTKNSPATRAVRLMEHEMKHVLVGLTKDLFGPRIRYRWADYFPTQSWELI
 YPKDNMLEVGGICIMRHEILORSRGVSGIAGVGLERLAWLFDPIRLFWNSDS
 GFLSQSEKDNHLPRKPKSHIPQCTNDLSFLLPQDIEVDAGFSNDPYDLYRSVAG
 DWVEQISLVDKERKPKTKGSSVCFRIVTRHMERLTIOAEVNEIHKQIASASVDSFVQ

BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
 ORIGIN
 Query Match 4.3%; Score 43; DB 21; Length 10772;
 Best Local Similarity 19.9%; Pred. No. 1.75e-06;
 Matches 32; Conservative 75; Mismatches 52; Indels 2; Gaps 2;

Db 1677 AGARAGTYTTAAKRWYAAWTTTMMKWWTTWKMAAMKTYTMMWKKMTSRITTT 1736
 || : : : || :
 Cp 2011 AGTAAATCGGAAATTTTGACCCCTGGAATTTTATCTGATTTCTGCGGA-ATTTT 1953
 :
 Db 1737 SAMMMWMTWTWKYMAWMMKMTWTWARMASWAPKMTSAAAYSA-RKMK 1795
 :
 Cp 1952 GGGGTTTTTTTGTAAATTTGTTTAAACAGTTCCTTCITTTTTCGCAACATRAA 1893
 :
 Db 1796 WMAVRAKRTWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 1836
 :
 Cp 1892 ACACAATTTTCAATGAGAAATCGTATTTTAAATTTGAAA 1852

RESULT 11
 LOCUS CEF35C5 33164 bp DNA INV 23-NOV-1998
 DEFINITION Caenorhabditis elegans cosmid F35C5, complete sequence.
 ACCESSION Z81076
 NID 91627897
 VERSION Z81076.1 GI:1627897
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditiida; Rhabditiida; Rhabditiida; Peloderiinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 33164)
 Sims, M.
 Direct Submission
 Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jee@sanger.ac.uk or rwenematode.wustl.edu
 2 (bases 1 to 33164)
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Beks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Riften, J., Roopra, A., Saunders, D., Shownkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkison-Sproat, J. and Wohldman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*
 Nature 368 (6466), 32-38 (1994)
 94150718

TITLE
 JOURNAL
 MEDLINE
 COMMENT

Coding sequences below are predicted from computer analysis, using predictions from GeneFINDER (P. Green, U. Washington), and other available information.
 For a graphical representation of this sequence and its analysis see: "
 http://webc.sanger.ac.uk/cgi-bin/display?db=wormacc&class=sequence&object=F35C5
 Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 IMPORTANT: This sequence is not the entire insert of clone F35C5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

Note: remainder of annotations omitted.

Query Match 4.4%; Score 44; DB 21; Length 41564;
Best Local Similarity 75.6%; Pred. No. 5,82e-07;
Matches 65; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Db 24994 GAAATAGTCGCAAGAATAAATATGGCTGCATTGTCTGTAATAATAATTTTTCGA 25053
|||||
Cp 2766 GAAAATTAGTGTTAAAAATAATAATACCTCGAGCTTCAGTAATAATAATAATATTCCA 2707
||| |

Db 25054 GCTGAACAATGCGAAAGTCACAA 25079
| ||| |

Cp 2706 CTTCGAATATTGGGAAAGTCATCAA 2681

gene /evidence=not_experimental
/protein_id="AACJ19217.1"
/db_xref="PID:g3193188"
/db_xref="GI:3193188"
/translation="MYEDQYCVGNFLEBGTGDPRDGLPTITASHATSMDMLILEKM
TSMMDGVISGIFIDFHSSOALEYLAEVHRCDEEPRKMTIHFAIROSAPFOOTCKIO
IPASRTCMKFRADDSYLSRLSHSGFOLYSUNLMLNAROGKSDIRIMADMIVSE
GEARKLVANEMTDGKSKKYLAIRFESSVNTYIPIRHFELKOSMANSNGTEWVOV
ILHRDPYNAAYFPFSRIKVMShLYALCRAGTIVPSPHVEDVHGIIKHNTITSKAT
IAHDPAAMDIAGARVREMDEKYPDTLDKGRFKMY"
19348..23506
/gene="F22f7.2"
join(19348..19506,19621..19740,20773..21033,21767..22174,
22868..23082,23401..23506)
/gene="F22f7.2"

CDS /evidence=not_experimental
/codon_start=1
/evidence=not_experimental
/protein_id="AACJ19222.1"
/db_xref="PID:g3193193"
/db_xref="GI:3193193"
/translation="WEARYDVVYGATVGTKGFETLLISCGEDFNYSIAIAGRSEK
LEEVLERSKTSTSLKTRIGLIVCDSTNESMGMARAKLIYAIVPFRIHGAAY
KAAENGOANDIVAGEPEMWIERMEAKYQOMAKNNVVITYSAGMSIPRHLKLR
YFDLHORIDAFLOLHPGSGSFSAATLOALLGFNPADIGSI.RSHIMTKRIDHTL
APNKRRHKLIEEGSGSAMPLPDGAKSYVTBSQYDYDMANNKVPRIFFTFES
RWNLILLAAIAGLKLPFSKTYTMQRFIRRYBELSCGLPSGSGBEEDMKATFTYO
YGYIGDPRDRPKRLVLTCTGDDPVYIATSGCLSVLFLKEKELPKSGVY
TTAAAFENTRLIDFLDLFGIKEDIEMPEKL"
24016..24017
/note="SL1 trans-splice site; see EST cm12d3"
24018..26348
/gene="F22f7.1"
join(24018..24183,24228..24334,24381..24801,25685..26165,
26243..26348)
/gene="F22f7.1"
/note="coded for by C. elegans cDNA YK139b8.3; coded for
by C. elegans cDNA cm14g4; coded for by C. elegans cDNA
YK53a9.5; coded for by C. elegans cDNA YK139b8.5; coded
for by C. elegans cDNA cm12g3; coded for by C. elegans
cDNA cm15e5"
/codon_start=1
/protein_id="AACJ19221.1"
/db_xref="PID:g3193192"
/db_xref="GI:3193192"
/translation="MSRDVVVYGASGFTGAVVEYILVNSDFEGLSPAAGRSEKTL
REVRLNISOKTGKDVSNAAYIVADSADERSLENAROANVINAQPYRLGEAVKA
AVEKGASHVIDSGEPAMTEKKMOQYKROAKRSGVYINAGPRLTGEAVKA
HGDINHVESFYOLLTGPBGYFPNACTYTOILLIGSAATDGLAKVRKINEKTYRGA
VKLPKRPTIMEIKRELKNVAVPFPGAKSLIINSQYIDAISROWRPIMHEITYRLSS
OFYVYLGWLIMPLSIFYKVFETRRIILOQYPDQCSFEWNKSGSARMABASEVYW
FGYGVKTLPLDQDHGKINKRVLATCGPDAGYIAISGVLSAALTIRDKDNLPKD
GGVYTTAAAFENGSKIYDIASFGITVQLSEYDL"
complement(28572..30790)
/gene="F22f7.7"
complement(join(28572..28680,29262..29452,29953..30069,
30620..30790))
/gene="F22f7.7"
/note="contains similarity to E. coli cation transport
protein (GB:L28709)"

RESULT	9		DNA	HTG	02-MAR-1999
LOCUS	CEY53F4	495882 bp			
DEFINITION	Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** FROM				
ACCESSION	clone y3f4, WORKING DRAFT SEQUENCE.				
NID	Z92860				
VERSION	g4464195	GI:4464195			
KEYWORDS	HMG; HTGS PHASPL				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditoidea;				
AUTHORS	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis				
TITLE	1 (bases 1 to 495882)				
JOURNAL	Smye,R. Direct Submission Submitted (02-MAR-1999) Nematoide Sequencing Project, Sanger Centre Hinxton, Cambridge CB10 1RO, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jesse@sanger.ac.uk or rtw@nematoide.wustl.edu On Mar 22, 1999 this sequence version replaced gi:4455382. Order of segments is not known: 800 n's separate segments. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that this sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. * NOTE: This is a 'working draft' sequence. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. Location/Qualifiers source location=82 1..495882 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="I" /cloned=Y53f4"				
FEATURES					
BASE COUNT	153420 a	87589 c	84294 g	153691 t	16888 others
ORIGIN					
Query Match	4.4%	Score 44:	DB 19:	Length 110000:	
Best Local Similarity	76.88:	Pred. No. 5.82e-07:			
Matches 63:	Conservative	0:	Mismatches 19:	Indels 0:	Gaps 0:
Db 56849 GCATTTTCAACAATCTTCACATAAATAAGTAGATTATTTCCAATTAATCAAGTGATTTT					56908
QY 2685 GCACTTCCCAAATATTTGAAGGAATATTTATTATTACGAAGAAGCGAGCATATT					2744
Db 56909 CTATTTTTGCACACTAATCTC					56930
QY 2745 TTATTTTAAACACTAATTTTC					2766
RESULT 10					
LOCUS	AF012089	10772 bp	DNA	INV	05-AUG-1997
DEFINITION	Drosophila melanogaster cysteine proteinase-1 (CP1) gene, complete cds,				
ACCESION	AF012089				
NID	g2305220				
VERSION	AF012089.1	GI:2305220			
KEYWORDS	.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryotae; Mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephyrroidea; Drosophilidae; Drosophila.				
AUTHORS	Gray,Y.H., Tanaka,M.M. and Syed,J.A.				
TITLE	p-element-induced recombination in Drosophila melanogaster: hybrid element insertion				
JOURNAL	Genetics 144 (4), 1601-1610(1996)				

QY 1969 CAAGATMAAAATTCAGAGGTCATAATTTCCGA 2004

LOCUS 8 CELF22F7 41564 bp DNA INV 09-JUN-1998
DEFINITION Caenorhabditis elegans cosmid F22F7.
ACCESSION AF067937
NID 93193187
VERSION AF067937.1 GI:3193187
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.

REFERENCE
AUTHORS
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fullon, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latrille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaubin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J., and Wohlman, P.

TITLE
2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*

JOURNAL
MEDLINE
94150718
2 (bases 1 to 41564)
2 (bases 1 to 41564)
AUTHORS
TITLE
The sequence of *C. elegans* cosmid F22F7
Unpublished (1998)
REFERENCE
AUTHORS
3 (bases 1 to 41564)
Waterston, R.
Direct Submission
Submitted (23-MAY-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
e-mail: r.w@nematoe.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F36F12, 200 bp overlap; 3' cosmid is C45H4, 1100 bp
overlap. Actual start of this cosmid is at base position 1 of
CELF22F7.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

1. 41564
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="F22F7"
/chromosome="Y"
complement(140..2558)
/gene="F22F7.4"
complement(join(140..434,596..1324,1393..1508,1561..1669,
1906..2037,2278..2349,2410..2558))
/gene="F22F7.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19218.1"
/db_xref="PDB:93193189"
/db_xref="GI:3193189"
/translation="MPEKPYNFILPRYRKHIPRLVLELAILLGLSIFVLR
FLPKVSKIOFPOSTHTFISAYYPIKSLGNVAVMTNKRITVSKITRTINLI
GNTSPWKATLTTEHALDDRCDSYLLAOTNLEMTTELEADGILMOIPKRP
KYSAPKPYEVCSQPAEOMOTLVOLHYSKRGALQYVSMVSFNLKEVEX
LGLVSIPEWLTKEPTDGPYLENRNVELNQAAHTDCLIMKEVSTGYDMD
ILIPITNANSYEEFEREYAGSWELALQYKFPRTIKSALEASISLTVKARLP
TKDQKCFEPRERENSSMSHSRNSDHAIVLQDQHRHMRKVGQANGIFHMKM
TMVLEHVPBGALFVNDGNTTELISPOHLEKEIDMKRYSVHMRMLTRPEIANSIL
PNDFVNIYVQCYNESSEFHRDNEFLYNDVTCVNAFNCCLPQRENIPCHSATYS
GPKMPTTHYHAINSPFSRDICTQ"
complement(2982..5782)
/gene="F22F7.5"
complement(join(2982..3092,3138..3278,3343..3490,
3719..3795,3841..4003,4131..4241,4289..4537,5564..5691,
5765..5782))
/gene="F22F7.5"
/note="contains similarity to the choline/ethanolamine
kinase family (SW:P35790)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19223.1"
/db_xref="PDB:93193194"
/db_xref="GI:3193194"
/translation="MISNIGTLFAENSLESDVILEKYLEGSDYLRGMAVLEKSEV
VTQITGGSNIIYATSNRTKLSIDPSCFELIRHQAPSQVFEIVVFSVLSRGK
PKIYGFPGGRLEELPSRTLDVDSIKLEIARQSGVLYPKYHIEVPIKSGALKF
IKENLEGKALGCVKLMCPHSVYSDMPSLETLYEOLIEATFERKCKVDEYVS
HNDLAPLVLELNTKEVETIDFESYNNRGPICMFLDEMDYDDPPLGIVDQ
DYMEHPVLQICEAVYIDTQYKMTAPKQFPLTEGRAAEVSLMECCFFPLVNM
FWAWSLROHLAKYEIGMDLVIASDRSMYFLKEQSRALYEQFK"
7223..12040
/gene="F22F7.3"
join(7223..7470,7904..8035,8081..8189,8246..8361,
9353..10066,10676..10947,11679..11779,11909..12040)
/gene="F22F7.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19219.1"
/db_xref="PDB:93193190"
/db_xref="GI:3193190"
/translation="WETFLHPIGYSLTLTKKYPRLHPRIKIAITLILCLMI
TPLFNDSFLEIDKVEVLIQSTHTFHSAYYIPVSKSLNNVAIVTNNKRTVPI
ITDYINIHGSHIKKKVKAVALTTEHLINRCDSYLVLIQANLGMGTKLEISGV
ITVEIPKSPKYSAPRVVFCISPOFAEOMOTLVOLHISKRGALHVLVYVESY
KLIEKEKLGIVSIEPWLTIKFPVTGSPYLENRNVELNQAAHTDCLIMYEAVS
FVGLIDMDITLIPNKANSYEEFEREGGSYISALQYKRADEFTIVAELEASISA
IVKNAKRLPTKDQKSFVPRERENSSMSHSRNSDNKPMWTPTQVPLSKRKAKOI
NGIFHMRMLVLTNLKRVGDGAVPNPQGNITQLISDHLAETIDEMKSLITSEIAL
SATLEKTPFVPIIFKCYNDSFYLHBDTKTLNVTVCVNAFDELPOOEGMPCYHSA
TYHSGPPEPITFHWATDAYSIGRQOVLVLRFGNDKCGYFGIRIRIPPRHNP
OPNSTCGLVHNSFTYDKRKFIILKVRSTGKMPKKE"
complement(12037..14830)
/gene="F22F7.6"
complement(join(12037..12118,12187..12323,12377..12470,
12794..13073,14219..14321,14521..14713,14778..14830))
/gene="F22F7.6"
/codon_start=1

ORGANISM	Caenorhabditis elegans
REFERENCE	Eukaryote: mitochondrial eukaryotes; Metazoa; Nematoda; Rhabditidae: Rhabditia: Rhabditida: Rhabditina: Rhabditoidea: Rhabditidae: Caenorhabditis.
AUTHORS	1 (bases 1 to 35713) Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hallier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Lalster, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Riken, L., Roopra, A., Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkison-Sproat, J. and Woldman, P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6466), 32-38 (1994)
REFERENCE	94150718
AUTHORS	2 (bases 1 to 35713) Leimbach, D.
JOURNAL	The sequence of C. elegans cosmid F22D3
REFERENCE	Unpublished (1995)
AUTHORS	3 (bases 1 to 35713) Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (12-JUN-1995) Robert Waterston
COMMENT	Submitted by: Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England e-mail: r.w@emsl.mrc.cam.ac.uk
NEIGHBORING COSMID INFORMATION:	The 5' cosmid is B0252, 200 bp overlap; 3' cosmid is C15F1. Actual start of this cosmid is at base position 1 of C15F22D3; actual end is at base position 35713 of C15F22D3.
NOTES:	Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).
FEATURES	Location/Qualifiers
source	1..35713 /organism="Caenorhabditis elegans" /strain="Bristol N2" /db_xref="taxon:6239" /map="11" complement(join(108..263,313..385,428..504,878..955)) /gene="F22D3.4" complement(join(108..263,313..385,428..504,878..955)) /gene="F22D3.4" /codon_start=1 /evidence=not_experimental /protein_id="AA68395.1" /db_xref="PIR:g861378" /db_xref="GI:861378" /translation="MNLVLLVLSSTFLCAAEIITYTNGNCKEKGSGDPAEAVENSDIPEVHNSKNSKIMELSKKEY" SPIDPEVHNSKNSKIMELSKKEY"
gene	join(1168..1258,2416..2438,2944..2965,3774..3794,4215..4251,4310..4410,5631..5710,5761..5833,5919..6003,6131..6179,6619..6796,7216..7282,7331..7405,7451..7583,8018..8140,9089..9163) /gene="F22D3.3"
CDS	join(1168..1258,2416..2438,2944..2965,3774..3794,4215..4251,4310..4410,5631..5710,5761..5833,5919..6003,6131..6179,6619..6796,7216..7282,7331..7405,7451..7583,8018..8140,9089..9163) /gene="F22D3.3"
ORGANISM	Caenorhabditis elegans
REFERENCE	Eukaryote: mitochondrial eukaryotes; Metazoa; Nematoda; Rhabditidae: Rhabditia: Rhabditida: Rhabditina: Rhabditoidea: Rhabditidae: Caenorhabditis.
AUTHORS	1 (bases 1 to 35713) Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hallier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Lalster, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Riken, L., Roopra, A., Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkison-Sproat, J. and Woldman, P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6466), 32-38 (1994)
REFERENCE	94150718
AUTHORS	2 (bases 1 to 35713) Leimbach, D.
JOURNAL	The sequence of C. elegans cosmid F22D3
REFERENCE	Unpublished (1995)
AUTHORS	3 (bases 1 to 35713) Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (12-JUN-1995) Robert Waterston
COMMENT	Submitted by: Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England e-mail: r.w@emsl.mrc.cam.ac.uk
NEIGHBORING COSMID INFORMATION:	The 5' cosmid is B0252, 200 bp overlap; 3' cosmid is C15F1. Actual start of this cosmid is at base position 1 of C15F22D3; actual end is at base position 35713 of C15F22D3.
NOTES:	Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).
FEATURES	Location/Qualifiers
source	1..35713 /organism="Caenorhabditis elegans" /strain="Bristol N2" /db_xref="taxon:6239" /map="11" complement(join(108..263,313..385,428..504,878..955)) /gene="F22D3.4" complement(join(108..263,313..385,428..504,878..955)) /gene="F22D3.4" /codon_start=1 /evidence=not_experimental /protein_id="AA68395.1" /db_xref="PIR:g861378" /db_xref="GI:861378" /translation="MNLVLLVLSSTFLCAAEIITYTNGNCKEKGSGDPAEAVENSDIPEVHNSKNSKIMELSKKEY" SPIDPEVHNSKNSKIMELSKKEY"
gene	join(1168..1258,2416..2438,2944..2965,3774..3794,4215..4251,4310..4410,5631..5710,5761..5833,5919..6003,6131..6179,6619..6796,7216..7282,7331..7405,7451..7583,8018..8140,9089..9163) /gene="F22D3.3"
CDS	join(1168..1258,2416..2438,2944..2965,3774..3794,4215..4251,4310..4410,5631..5710,5761..5833,5919..6003,6131..6179,6619..6796,7216..7282,7331..7405,7451..7583,8018..8140,9089..9163) /gene="F22D3.3"

6498..6602,6658..6748,6791..7021)
 /gene="C16A11.5"
 /evidence=not_experimental
 /codon_start=1
 /protein_id="AAC26261.1"
 /db_xref="GI:3319386"
 /translation="MCSINELITGKKEIPEDPLTASRLMDAADYVPLKRLGHIDAND
 INSSPKTLENGSTFEFFOALIDTHPRLLYSGKELANVRIHNRFPISHRSPITYS
 LDNOYITKIDFILLIONICGRKSTNITIELNLEMYLAKAESEKRYLEFVE
 DADSRKIEKFEKFNQACQVDRSLERONSNDHLPIVOLDGMSKIEFTIOHIOA
 KRAVPIEDLDGSGYCIILAVDALEFNEIINGSVFEKHKVPKIIIEVYLIPILLRK
 SELRPHPIKTAHRLHRLASEVSLSPHPANPYKEVRNAKADHGFTIOLKMEITY
 LGTEFPDITDYAETIAYNEVRYKEMLLRCDFPIAIEHCQLCTFNPSPIKQLE
 HNOKACHRLVGHODICIKVEKSENEANDLDESSSTKSPIDCKCSOTLQTIEDV
 KKELTSONKEMKMKRYSELEKLDVNOQEVEMEDDLGQEBNLITANDSL
 KYHDVLMELIADIKONHSLKLEHVEGLKEENIKOERNYNTLOSNEANLSREGLIMC
 GFSPTYEELDQEROKYOMLTARSEETSYQAEPAAGGEEMLDVTQIIEKKEEKIRL
 KAEHDANERIIQOOLLDKLANFS"
 7825..9790
 /gene="C16A11.4"
 join(7825..8408,8461..9120,9306..9678,9728..9790)
 /gene="C16A11.4"
 /note="coded for by C. elegans cDNA YK334C12.3; coded for
 by C. elegans cDNA YK298a8.3; coded for by C. elegans cDNA
 YK103h8.3; coded for by C. elegans cDNA YK151d2.3; coded
 for by C. elegans cDNA YK151d2.5; coded for by C. elegans
 cDNA YK103h8.5; coded for by C. elegans cDNA YK298a8.5;
 coded for by C. elegans cDNA Cml7b3"
 /codon_start=1
 /protein_id="AAC26260.1"
 /db_xref="PID:g3319385"
 /db_xref="GI:3319385"
 /translation="MDDSEBANOFEPPEPRPAIKRSTAVNNLSPHQKVCQCCTNTND
 NMLIKVCCEKMOAICMSVPEIDITFTICEMCEPRLLAVSSQAAYGVGPEPRE
 VRKGRNOSVLEEFLELSEASNISRLPALTITLEDYKEINIPQSVAILKROLES
 NLEETISPMVYADRAQOMLFAMGLVKEGFELEERNAVEVDGHRITFEKSNL
 EFRGHRVGRPLVNRKRSASETSQOHSNGEEDDDIFIDLVPVATGSRNMQPFVK
 QEIPDEQSGASITLPNAETEPNLOYSIDMTTGSATWNMAATPSEBRGIMTSM
 NGHALOSASPYQEOEMICVOHMTKMSMHTIKQSPFYEMKTRFGMKINITS
 RFSYHFFPGCYLAFELANSFLIAAPAPAPATITLRAFLHFKGVIVLESEQLVD
 CSRTQOHLCTCRADKLISMEKLTILAVATTAELHSDVAIINGASAMFEYKEFI
 EITVSNOSNSAFSGLCITDLKQIALVNMCKIIPFKELSRAPRTAISLSVQ"
 10319..15091
 /gene="C16A11.3"
 join(10319..10881,10929..11022,11160..11339,11392..11442,
 12018..12297,12344..12408,12962..13039,13096..13374,
 13959..14338,14659..14743,14966..15091)
 /gene="C16A11.3"
 /note="similar to protein kinases (Pfam: pkinase.hmm,
 score: 168.47)"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAC26259.1"
 /db_xref="PID:g3319384"
 /db_xref="GI:3319384"
 /translation="MEHDSSEDSDOFEPPRPVYSIKIECQCGTNTNRLIKVCYCA
 QMHAOCMVGRPTDVCYAYKCEDCEPRLLITVIAEESQVGVPEPPTAPRIERTT
 KYLYOFLELDLSDKSENKSPLETKFPQDYTKSGSGSLATARKKILLAEITV
 MSAYNDNRKAEMLPAIGLEVEDEFLERGRATVVDCCORTTKSTSTSSPEV
 SPVLAATTAIAEHALDSAYIDGISAIEFKYIENYAKRALISLDDLOKIQIGLEA
 MDKKIYNVTSRVFOATISLRKVOKLITGEGMSKNGNLMARKFNQOFSKMFVPL
 KTKTFKQMTKRAPEONDEIKIRRKSETDENVAQRNPKNPILITSSGAOLKI
 IAKIGEVGAAYKVVDHSNKYAAVLFEBHESEVGEQMAETALMKAVTAIANSAL
 IRMEDSRINESVNGFLKAIISFELKGRSVFEMDSTGKMPAGSEISFEELANRYM
 GKOLLEMEKLESVKIVLHDLKPEKICFSSSCFCTTGVACFIOPSTFHCYVDFG
 NARTIRQDPRKVALVOTONYRAPEITELIPKVSCTTGVSCCITSELTGGLTFTYGN
 SKSDTELEIEMOMTVQOPPCAMREAKIKSTRIRIENGWVTKRKSSEFPBK
 PLKQRKRLLELIPFLDLECLIVDPSSRRLDISRTFFK"
 complement(15718..18718)
 /gene="C16A11.6"
 complement(join(15718..16143,16238..16444,16498..16633,
 16683..16762,17361..18038,18257..18392,18438..18517,

18581..18718))
 /gene="C16A11.6"
 /evidence=not_experimental
 /codon_start=1
 /protein_id="AAC26262.1"
 /db_xref="PID:g3319387"
 /db_xref="GI:3319387"
 /translation="MRNAKEMHADITVNLKGTDTLLHGRSDRVNRSIFQFLFRI
 LITSAMNSNPITVPSMOCITIKHLANKRFOIYVACQSLYKIEKQPLIKELSKRR
 TISVNGVEKIKAKRFEAPHTLAIKOLGRQPNVWDEPDDGRILDELSFKK
 KYNPNEDNSDNPILDRFDLMSPELKVLSDRSLNLDVIDHEILRLRSEGLMPH
 IYPRELPHGILSTELIVANEVGEVDFLWIDALLQPDGNGADSFITIEDDLI
 ANLEKAFETLLAAGKDRALSEYEGDEDECFRRARYTCDEPNNGIWFNDYVIA
 DAKTEKNTNISAKMNSNPITYPSMOCITIKHLANKRFOIYVACQSLYKIEKQPLI
 KKLIFNRTISFNGVYKYGAKNKEAPHTLAIKOLGRQPNVWDEPDDGRILE
 LPSEKFEVQVHQMTRSSNDPLIERPIQKSLPKNGCSPPSGTIPALIRAGLNDIA
 EYDFPNTLIDALLQDREYTGCFITTEDVDVSELPEAFITLLAEKDAIYIEEDEN
 PDEFDGTYRCDEFFELIAFWENDDMVISADVKEKGDGFQIWLIREDEKTLIDLEFIN
 Y"
 18981..25463
 /gene="C16A11.2"
 join(18981..19013,19054..19152,19203..19306,20074..20557,
 21287..21579,22919..23371,24089..24284,25173..25463)
 /gene="C16A11.2"
 /note="coded for by C. elegans cDNA YK275g3.5"
 /codon_start=1
 /protein_id="AAC26258.1"
 /db_xref="PID:g3319383"
 /db_xref="GI:3319383"
 /translation="MDLNAAAAEKPYETSSCRNDITIMLSASGYHRAVSLNDYDK
 IYGGIYVAMRCDDVDVAVDLIYSESEONIKLRTEOSEKLEALKAVOCPEHATQ
 ISGIDFLAKNYIOMLRHVALQESKGRKFHHFPESMHHEHDEALQDRHLRLRG
 VTTTRYMRKFNDSAMFNLSMDAKCTIAEYVYSTRGDVGFTEDETDLKVMKSP
 ILSEKLEKIREFNFLNKRFEENASTKKHVPTVYRKMDQAVVETPNDESIKTSR
 EGMCEVEDELINPORLEALLESIVIDEALFGVVDTHKQSDRANTIKRNVSDQH
 TPEVNSLDLOAEIYKPEYTOARADLRACACQDVMAONHENLRLKONDKKMQ
 KYVCQRETVIDHNVATITDLITFSALIKGMDVEKSLVVEPQSOEYRQAFVSYNDVET
 LQGEYHMAKLTODKTLAEKDTLALRIKQKREVSNTDEMKRLKHFPEVKLNC
 KLPEYRLGNLIEIFFEINSGNSKRSKSPKISKKFQWLKYESLKEVDSLLAQR
 EISPAFOYALDSREAEDEDEAERS"
 33481..36049
 /gene="C16A11.1"
 join(33481..33632,33795..34271,34988..35116,35232..35274,
 35382..36049)
 /gene="C16A11.1"
 /note="similar to the G-alpha family of guanine
 nucleotide-binding proteins (Pfam: G-alpha.hmm, score:
 410.24)"
 ...
 Note: remainder of annotations omitted.
 Query Match 4.8%; Score 48; DB 21; Length 47147;
 Best Local Similarity 79.3%; Pred. No. 6,666-09;
 Matches 65; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 Db 24724 GAAATTAGTCTTAAATATAAATTAATCTGATTTATCGAATAATAAGTATTTC 24783
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 2766 GAAATTAGTCTTAAATATAAATTAATCTGATTTATCGAATAATAATATTTC 2707
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 24784 GCTGAAATTCGAGAAAGTGC 24805
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 2706 CTTCAATATATTGGGAAAGTGC 2685
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 6 CELF22D3 35713 bp DNA INV 15-JUN-1995
 LOCUS Caenorhabditis elegans cosmid F22D3.
 DEFINITION U28993
 ACCESSION U28993
 NID 9861377
 VERSION U28993.1 GI:861377
 KEYWORDS
 SOURCE Caenorhabditis elegans strain-Bristol N2.

CP 2560 TCAAGAACTTCAGCAAGTCCTTGGTCCGTAGAGGAGACATATATAACGGCTCG 2501

Db 10855 AACCCACATTCCTCCGCTGCTGACGTCTTTCAGATCTCCGTCCTCTCCGCAAC 10914

CP 2500 AACCCACATTCCTCCGCTGCTGACGTCTTTCAGATCTCCGTCCTCTCCGCAAC 2441

Db 10915 GTTCCACATGACCTCAAAATATCTTATATATCAATTTTAAATATATTC 10974

CP 2440 GTTCCACATGACCTCAAAATATCTTATATATCAATTTTAAATATATTC 2381

Db 10975 GATTAAAACTCACATTAATCATATCTCATTAATCACTATTAACAACCTGTTGGGATG 11034

CP 2380 GATTAAAACTCACATTAATCATATCTCATTAATCACTATTAACAACCTGTTGGGATG 2321

Db 11035 AGAAGCTCGGAATTCATGACCTTTAGATGACTAGAGAACATCAATATGTTCTCTCT 11094

CP 2320 AGAAGCTCGGAATTCATGACCTTTAGATGACTAGAGAACATCAATATGTTCTCTCT 2261

Db 11095 AGCAAGCTCTTATCTATCTGACCATCATCTTGGCTGATGATGATGCTCTACTGACC 11154

CP 2260 AGCAAGCTCTTATCTATCTGACCATCATCTTGGCTGATGATGATGCTCTACTGACC 2201

Db 11155 CGTAGCAGAGTCAGATTTGATTTTGGATTTAGTCTGAAATGAGTAATATTGGAAA 11214

CP 2200 CGTAGCAGAGTCAGATTTGATTTTGGATTTAGTCTGAAATGAGTAATATTGGAAA 2141

Db 11215 AATAGTGTATTTTAAACATGTCCTCAAAATCAGCTCATTTGAGTAATTTTATC 11274

CP 2140 AATAGTGTATTTTAAACATGTCCTCAAAATCAGCTCATTTGAGTAATTTTATC 2081

Db 11275 GATTTTTCCTAGTTTTCGTCAAAAACCTCAACAACCTGATGAACTGCAAAAAAAG 11334

CP 2080 GATTTTTCCTAGTTTTCGTCAAAAACCTCAACAACCTGATGAACTGCAAAAAAAG 2021

Db 11335 GTGAAGTCAGTAAATGGAATAATTTTACCCCTCTGATTTTATCTGATTTGCG 11394

CP 2020 GTGAAGTCAGTAAATGGAATAATTTTACCCCTCTGATTTTATCTGATTTGCG 1961

Db 11395 CGAATTTTGGGGTTTTTTTGTAAATGTTTGAATCACTCTCTTATTTTGGCA 11454

CP 1960 CGAATTTTGGGGTTTTTTTGTAAATGTTTGAATCACTCTCTTATTTTGGCA 1901

Db 11455 AACATAAACACAAATTTTCAATGAGAAATCGTATTTTAAATTTGAAACGATTTTAA 11514

CP 1900 AACATAAACACAAATTTTCAATGAGAAATCGTATTTTAAATTTGAAACGATTTTAA 1841

Db 11515 AATGTTTGTGTTTGTGTTTGTGTTGTCAGCTCGAA 11554

CP 1840 AATGTTTGTGTTTGTGTTTGTGTTGTCAGCTCGAA 1801

RESULT 3
LOCUS CEY67H2 314495 bp DNA HTG 04-MAR-1999
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y67H2, WORKING DRAFT SEQUENCE.
ACCESSION AL022475
NTD 94469034
VERSION AL022475.3 GI:4469034
KEYWORDS HTG; PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryote; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitida;
Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 314495)

REFERENCE
AUTHORS McMurtry, A.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridgeshire CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jamesmanger@wustl.edu or twenematode@wustl.edu
COMMENT On Mar 22, 1999 this sequence version replaced gi:4468145.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

Source
Location/Qualifiers
1..314495
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y67H2"
BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others
ORIGIN

Query Match 99.1%; Score 991; DB 19; Length 314495;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 240213 TTGGGCAATCTACAAAATGCGCTTTTGGCCACGAAATTTAGTGTAAATAATAATA 240272

CP 2800 TTGGGCAATCTACAAAATGCGCTTTTGGCCACGAAATTTAGTGTAAATAATAATA 2741

Db 240273 TCACGAGCTTTTCAATTAATTAATATATTTTCACTTCAATATTTGGGAAATGCAATA 240332

CP 2740 TCACGAGCTTTTCAATTAATTAATATATTTTCACTTCAATATTTGGGAAATGCAATA 2681

Db 240333 TTTCGATTTTTCGAGGAAAAAACCCTGCGGTGGGTTGCAAAATTTAGCATCGCGCTCT 240392

CP 2680 TTTCGATTTTTCGAGGAAAAAACCCTGCGGTGGGTTGCAAAATTTAGCATCGCGCTCT 2621

Db 240393 TTT-CCGTGAGCCCTTGCCTTTTGGCCGACCTTTTAAACCTACATCTGGCGAGAGT 240451

CP 2620 TTTTCGCTGAGCCCTTGCCTTTTGGCCGACCTTTTAAACCTACATCTGGCGAGAGT 2561

Db 240452 TCAAGAACTTCAGCAAGTCCTTGTGTCCTGAGAGCAAGACATCAATAAACGCGTGC 240511

CP 2560 TCAAGAACTTCAGCAAGTCCTTGTGTCCTGAGAGCAAGACATCAATAAACGCGTGC 2501

Db 240512 AAGGCCACATCTCCCGCTGTTGACATGCTTTCAGCATCTCCGCTCTTCGCGGAAC 240571

CP 2500 AAGGCCACATCTCCCGCTGTTGACATGCTTTCAGCATCTCCGCTCTTCGCGGAAC 2441

Db 240572 GTTCCACATGACGCAAAATATCTTTTATTAATCAATTTTAAATATATATTC 240631

CP 2440 GTTCCACATGACGCAAAATATCTTTTATTAATCAATTTTAAATATATATTC 2381

Db 240632 GATTAAAACTCACATTAATCATATCTCATTAATCACTATTAACAACCTGTTGGGATG 240691

CP 2380 GATTAAAACTCACATTAATCATATCTCATTAATCACTATTAACAACCTGTTGGGATG 2321

Db 240692 AGAAGCTCGGAATTTTCAATGACCTTTAGATGACTAGAGAACATCAATATGTTCTCTCT 240751

CP 2320 AGAAGCTCGGAATTTTCAATGACCTTTAGATGACTAGAGAACATCAATATGTTCTCTCT 2261

Db 240752 AGCAAGCTCTTATCTATCTGACCATCATCTTGGCTGATGATGATGCTCTACTGACC 240811

CP 2260 AGCAAGCTCTTATCTATCTGACCATCATCTTGGCTGATGATGATGCTCTACTGACC 2201

Db 240812 CGTAGCAGAGTCAGATTTGATTTTGGATTTAGTCTGAAATGAGTGAATTTATGGAAA 240871

CP 2200 CGTAGCAGAGTCAGATTTGATTTTGGATTTAGTCTGAAATGAGTGAATTTATGGAAA 2141

Db 240872 AATAGTGTATTTTAAACATTTGTCGCAAAATTTACACTGATTTAGTAATTTTATC 240931

CP 2140 AATAGTGTATTTTAAACATTTGTCGCAAAATTTACACTGATTTAGTAATTTTATC 2081

Db 240932 GATTTTTCCTAGTTTTCGTCAAAAACCTCCAGACATGCACTAGACGAAAAAAG 240991

CP 2080 GATTTTTCCTAGTTTTCGTCAAAAACCTCCAGACATGCACTAGACGAAAAAAG 2021

REFERENCE
AUTHORS

2 (bases 1 to 39908)
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Spoot, J., and Woldman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*

TITLE

Nature 368 (6466), 32-38 (1994)

JOURNAL
MEDLINE
COMMENT

Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
 For a graphical representation of this sequence and its analysis see: -

http://webpage.sanger.ac.uk/cgi-bin/displaydb-wormacc/class=Sequence&object=C48D1
 Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 This sequence is the entire insert of clone C48D1. The true right end of clone F58D2 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence 281093.
 The end of this sequence (33020..39908) overlaps with the start of sequence 282274.

FEATURES

source

Location/Qualifiers
 1..39908
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="C48D1"
 /complement(94..4439)
 /gene="C48D1.1"
 /complement(join(94..1020,1345..1951,2283..2580,2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
 /note="similar to C. elegans ZYG-11 like protein; CDNA EST EMBL:Z14909 comes from this gene; CDNA EST EMBL:R01768 comes from this gene"

gene

CDS

/codon_start=1
 /protein_id="CAB02846.1"
 /db_xref="PID:e1344741"
 /db_xref="PID:g3875048"
 /db_xref="GI:3875048"
 /db_xref="SPTREMBL:O02227"
 /translation="MTKPLHDIAECKIAEFQNSYDREPELDEKSNIVSYIMNRDKPFLNLTIKLNTKVTLLHPTLDENTARLQKPSLEKLEIGNSKSKWCIFPKRETHADVIVLEAVLENEETRNKLOYLIEKDKLKRMTAKIGNLPSLTLTMSDVHINEDDYFOLGNSNLSLDISGTNLINAGISRLKQLEVALTRDPSLATYTELKLDENLKHLRVIVSOTRFLNHSRNNIVVEQFTCKMEDELPRLDRTSLSGDILEELKTIISIDPNISVTEAEITSOYNDVIRNTATYHRSVOLVLCIARKFEFGDYCKAKARIISBDNLONLTPSCIKTIYKSMQTPDISIVTONGKICITIPMTXSNLQNGVTLVSEALMSEVYILKRMISBAYKVNISYNAVEKLVNKLGLDLEWGSQKNEKILSEFEIKALRTASSVLCOCIRHMSDELKSLSEKTEENVNKLVDLEWVGQKNEKILSEFEITIKTEKSSKACTNMIOAGIISISLSCKEKEVSDITTKCELLTNLSALNDSIOELYSSNVDKIGOANVSTER"
 complement(6258..11120)
 /gene="C48D1.2"
 /complement(join(6258..6349,6876..7065,7420..7547,7602..7815,8728..9059,10248..10501,10780..10925,10989..11120))

gene

CDS

/gene="C48D1.2"
 /note="similar to Caspase recruitment domain, ICE-like protease (caspase) p20 domain. ICE-like protease from this gene; CDNA EST YK19810.5 comes from this gene; CDNA EST YK49145.5 comes from this gene"
 /codon_start=1
 /protein_id="CAB02848.1"
 /db_xref="PID:e1344743"
 /db_xref="PID:g3875050"
 /db_xref="GI:3875050"
 /db_xref="SPTREMBL:O02229"
 /translation="MRQDRSLERINIMFSSHLKVDLEILEVIAKOVLSNDGMDINSGCYREKREIVKAVORGVADFAEDALRSGHGLAEVLEPLARVDSNAVEECMSRASHRSRAISPAQYSPTRVHRVSYSVSTODIYRARSRSRLSHSDHNHSSPPVNAFPSPQSSANSPTGCSGLSSSSRRSPESKAGPTOYIFHEEDNFVAPITISRYEDERTMYRNSSPRACLTINNEHEQMPRTNGTAAKDNLTNIFRCGVYVICDNLITGRGMLITRDFAKSHSDSAILVLISHGEENITIGDDIPISHEIYDLNMAANARLANRKRIYVQACRGERNDGEPVLDSVGPALFRGMNDRSDPLNPLGCVAPVOQVYVRRKPSQADILIAVYATTAQVYSRNSARSGSIQAVCEVSTHARDMVELTPEVKKVAVDIPPAQKIVLLAGSTRKLCRKTRDLSLPN"
 complement(16940..21616)
 /gene="C48D1.3"
 /complement(join(16940..17126,18419..18597,18985..19442,19553..19879,19925..19961,20333..20481,20789..20987,21037..21152,21373..21616))
 /gene="C48D1.3"
 /note="similarity to Salmonella sodium/proline symporter (SW:POTP_SALTY). CDNA EST EMBL:C13840 comes from this gene; CDNA EST EMBL:C11552 comes from this gene"
 /codon_start=1
 /protein_id="CAB02847.1"
 /db_xref="PID:e1344742"
 /db_xref="PID:g3875049"
 /db_xref="GI:3875049"
 /db_xref="SPTREMBL:O02228"
 /translation="MADLLGIYAIVFYLLIVGIMAGRSKSKSENGATEEVMLAGRNIGTLVIGFTMTATWGAIVNGIAELNGLIGLGOAVYGAISLYMGILFAKKMREGYITMDLPDFQWNELELIFGTFNFKRLGFLKLOTIIEIDPFQKYGORIGGLMYVPALEGFMTAALISALGATSVILGIDNNAVTLASGLAIVPYTFGGYVAYAVDVVGLFCIFVGLLILGLYVONRRNREKESLMDCLMLFAGIPVOYFQVLSKTRHAGQCTSPVAGVCIIMAIIPALICAIKRNTPMDTSDVSPNNNGIKVIRPDKRNMVPLPLFOYLIRNVATFGLGAAVAAVMSADSSVLSAASPMHNKILIRPHASEKEVILYMRALICVGMATIMATLITOSIIGMLYLCADIVYVITRPLLQVYMPKSNYTGSLAGYAVGLRLIGEPVLSLPEFIPYDGVLPFPPTITAMSSNATYVISOSEKLEFKSGRLSPEDWVGCVVNPIDHVPVLSQVSAVSSETLNMKVECDGM OFPOLQREHLQYIRTSRLMKIHYVITHRTKVIIPQAIKRDDELVPRVHLISQIFCIPSPR"
 BASE COUNT 13182 a 6945 c 7146 g 12635 t
 ORIGIN
 Query Match 99.3%; Score 993; DB 21; Length 39908;
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;
 Matches 999; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

/protein_id="AAA27982.1"
/db_xref="PID:9456417"
/db_xref="GI:456417"
/translation="MMKODRSLLEKRNIMFSSHLKVEILEVLIKAOVLSNDNMI
/ncbi_taxonomy="Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Caenorhabditis"
/ecbpsspshrsrslspagyspdrvhsdsvssstysodiyssarsrslshs
/sdrnyspspvnafpsopssansstgcsstgssrnsfksagpdyifhedm
/fvdpftrsvfdektmrnssprgcllennhefomptngttraddnlnfrcm
/gtyvickdnltrgmlltfdpakhshsgdlsallvlsghenvitigvndpistheli
/ydlanapnrlankprifvqacrgerrndgfpvldsvdygafpflrrgmndgplf
/nflegcprvovwrrkksqadiliraytaoyvswrnarsgswfioavcevfsthak
/mdvvellevnkvacgfpotsognlkkompmtsrllkkyfwpfearnsav"
intron
2367..2429
/gene="ced-3"
/number=1
/exon
2430..2575
/gene="ced-3"
/number=2
intron
2576..2853
/gene="ced-3"
/number=2
exon
2854..3107
/gene="ced-3"
/number=3
intron
3108..4302
/gene="ced-3"
/number=3
exon
4303..4634
/gene="ced-3"
/number=4
intron
4635..5546
/gene="ced-3"
/number=4
exon
5547..5760
/gene="ced-3"
/number=5
intron
5761..5814
/gene="ced-3"
/number=5
exon
5815..5942
/gene="ced-3"
/number=6
intron
5943..6297
/gene="ced-3"
/number=6
exon
6298..6537
/gene="ced-3"
/number=7
intron
6538..7012
/gene="ced-3"
/number=7
exon
7013..7652
/gene="ced-3"
/number=8
BASE COUNT 2429 a 1455 c 1271 g 2498 t
ORIGIN
Query Match 100.0%; Score 1000; DB 21; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 1981 TTCAGAGGGCTCAAAATTTTCCGATTTTACGACTTTCACCTTTTTCGAGTTCAGT 2040
Qy 1981 TTCAGAGGGCTCAAAATTTTCCGATTTTACGACTTTCACCTTTTTCGAGTTCAGT 2040
Db 2041 GCAATGTTGGAGTTTTCGAGAACTAGGAAAAAATCGATAAAATTTACTCAATCG 2100
Qy 2041 GCAATGTTGGAGTTTTCGAGAACTAGGAAAAAATCGATAAAATTTACTCAATCG 2100
Db 2101 AGCTGAATTTTGGAGACAAATGTTTAAAAAACAATTTTTCATTAATTTTCACTCAT 2160
Qy 2101 AGCTGAATTTTGGAGACAAATGTTTAAAAAACAATTTTTCATTAATTTTCACTCAT 2160
Db 2161 TTTGAGCTTAATCGAAATTCGATCTGCTGACGAGGCTGCTAGAGAGCTCAACC 2220
Qy 2161 TTTGAGCTTAATCGAAATTCGATCTGCTGACGAGGCTGCTAGAGAGCTCAACC 2220
Db 2221 ATCAGCCGAATGATGCTGCTGACGATGAGAGAGCTGCTGAGAGAGCAATTTGATGT 2280
Qy 2221 ATCAGCCGAATGATGCTGCTGACGATGAGAGAGCTGCTGAGAGAGCAATTTGATGT 2280
Db 2281 TCTCTAGTATCTAAAGTGCATGAATTTCTCGAAGTCTTCATCGCAAAACAGTGTGA 2340
Qy 2281 TCTCTAGTATCTAAAGTGCATGAATTTCTCGAAGTCTTCATCGCAAAACAGTGTGA 2340
Db 2341 ATAGGATATATGAAGATATTTTTCGAGTATGCAAGGTTCCGAGAGAGAGCGG 2400
Qy 2341 ATAGGATATATGAAGATATTTTTCGAGTATGCAAGGTTCCGAGAGAGAGCGG 2400
Db 2401 AATTGATATATGAAGATATTTTTCGAGTATGCAAGGTTCCGAGAGAGAGCGG 2460
Qy 2401 AATTGATATATGAAGATATTTTTCGAGTATGCAAGGTTCCGAGAGAGAGCGG 2460
Db 2461 AGATCGTGAACACATGCAAGAGCGGGAGATGCGGCTTGCAAGCTTTATGATGCTC 2520
Qy 2461 AGATCGTGAACACATGCAAGAGCGGGAGATGCGGCTTGCAAGCTTTATGATGCTC 2520
Db 2521 TTGCGTCTACGGGACACGAGGACTTGTGAAGTCTTGAACCTCTCCAGATCGTAGG 2580
Qy 2521 TTGCGTCTACGGGACACGAGGACTTGTGAAGTCTTGAACCTCTCCAGATCGTAGG 2580
Db 2581 TTTTAAAGTTGGCGCAAAAGCAAGGCTTCACGAGAAAGAGGGGAGTATTTT 2640
Qy 2581 TTTTAAAGTTGGCGCAAAAGCAAGGCTTCACGAGAAAGAGGGGAGTATTTT 2640
Db 2641 GCAACCCACCGGACGAGTTTTCCTCCGAAATGGAATTTATGCACTTCCCAATAT 2700
Qy 2641 GCAACCCACCGGACGAGTTTTCCTCCGAAATGGAATTTATGCACTTCCCAATAT 2700
Db 2701 TTGAAGTGAATATATTTTATTTACTGAAAGCTGAGATTTATTTTAAACACTA 2760
Qy 2701 TTGAAGTGAATATATTTTATTTACTGAAAGCTGAGATTTATTTTAAACACTA 2760
Db 2761 ATTTTCGTGGCGCAAAAGCGATTTGTGATTTGCGGCA 2800
Qy 2761 ATTTTCGTGGCGCAAAAGCGATTTGTGATTTGCGGCA 2800

```

RESULT 2
LOCUS C48D1 39908 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION Z81049
NID 91627677
VERSION Z81049.1 GI:1627677
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORNTISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 39908)
AUTHORS Burton, J.
TITLE Direct Submission
JOORNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jesusanger.ac.uk or rw@nematode.wustl.edu

 M O S E R

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 03:10:58 1999; Msrch time 1828.29 Seconds

Tabular output not generated. 1516.122 Million cell updates/sec

Title: >US-08-287-669-18
 Description: (1801-2800) from US08287669.seq (3 of 10)
 Perfect Score: 1000
 N.A. Sequence: 1801 TTCACAGCTGACACACAGAA.....CATTTGAGATTGCCGAA 2800
 Comp: AAGGTCCGACCTGTTGCTT.....GTAAACATCTTAACGGCTT

Scoring table: TABLE default

Match STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

emb158
 1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
 7:em_in 8:em_cm 9:em_or 10:em_ov 11:em_pat 12:em_ph
 13:em_pl 14:em_to 15:em_sts 16:em_v1
 genbank111
 17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
 22:gb_in2 23:gb_cm 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
 28:gb_pl2 29:gb_pl 30:gb_pr1 31:gb_pr2 32:gb_pr3 33:gb_pr4
 33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 11.320; Variance 7.832; scale 1.445

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	100.0	100.0	7653	21	CELCED3A	0.00e+00
2	99.3	99.3	39908	21	CEC48D1	0.00e+00
3	99.1	99.1	314495	19	CEY79H2	0.00e+00
4	76	7.6	7218	25	CEY67H2	0.00e+00
5	48	4.8	47147	21	CELC16A11	1.82e-23
6	45	4.5	35713	21	CELF22D3	6.66e-09
7	44	4.4	10772	21	AF012089	1.92e-07
8	44	4.4	41564	21	CELF22F7	5.82e-07
9	44	4.4	110000	19	CEY35F4	5.82e-07
10	43	4.3	10772	21	AF012089	5.82e-07
11	43	4.3	33164	21	CEY35C5	1.75e-06
12	43	4.3	315330	19	CEY46G5	1.75e-06
13	42	4.2	26276	21	CELF53E10	5.20e-06

14	42	4.2	145920	19	AC006710	Caenorhabditis elegans	5.20e-06
15	42	4.2	224746	19	CEY56A3	Caenorhabditis elegans	5.20e-06
16	42	4.2	249640	19	CEY79H2	Caenorhabditis elegans	5.20e-06
17	42	4.2	229919	19	AC006712	Caenorhabditis elegans	5.20e-06
18	41	4.1	22881	21	CEY2C518	Caenorhabditis elegans	1.54e-05
19	41	4.1	39973	21	CEM106	Caenorhabditis elegans	1.54e-05
20	41	4.1	39973	21	CEM106	Caenorhabditis elegans	1.54e-05
21	41	4.1	74371	31	AC005369	Homo sapiens chromosome	1.54e-05
22	41	4.1	110000	19	CEY11B2	Caenorhabditis elegans	1.54e-05
23	41	4.1	125590	21	CEY19E10	Caenorhabditis elegans	1.54e-05
24	41	4.1	179854	19	CEY19D2	Caenorhabditis elegans	1.54e-05
25	41	4.1	179854	19	CEY19D2	Caenorhabditis elegans	1.54e-05
26	41	4.1	260699	20	AC006893	Caenorhabditis elegans	1.54e-05
27	40	4.0	38166	21	CEY43C9	Caenorhabditis elegans	4.50e-05
28	40	4.0	145920	19	AC006712	Caenorhabditis elegans	4.50e-05
29	40	4.0	29919	19	AC006710	Caenorhabditis elegans	4.50e-05
30	39	3.9	11230	21	CEY102A5A	Caenorhabditis elegans	1.30e-04
31	39	3.9	25600	21	CEY120F5	Caenorhabditis elegans	1.30e-04
32	39	3.9	29249	21	CEY18D11	Caenorhabditis elegans	1.30e-04
33	39	3.9	36355	21	CEM01F1	Caenorhabditis elegans	1.30e-04
34	39	3.9	42614	21	CEY0513	Caenorhabditis elegans	1.30e-04
35	39	3.9	75073	21	CEY48A6B	Caenorhabditis elegans	1.30e-04
36	39	3.9	110000	19	CEY102A5	Caenorhabditis elegans	1.30e-04
37	39	3.9	224746	19	CEY102A5	Caenorhabditis elegans	1.30e-04
38	39	3.9	244513	19	CEY79H2	Caenorhabditis elegans	1.30e-04
39	39	3.9	249640	19	CEY79H2	Caenorhabditis elegans	1.30e-04
40	39	3.9	261002	19	CEY41C4	Caenorhabditis elegans	1.30e-04
41	39	3.9	296699	19	CEY48A6	Caenorhabditis elegans	1.30e-04
42	39	3.9	298406	21	CEY75B8A	Caenorhabditis elegans	1.30e-04
43	39	3.9	316170	19	CEY37A1	Caenorhabditis elegans	1.30e-04
44	39	3.9	336538	19	CEY75B8	Caenorhabditis elegans	1.30e-04
45	38	3.8	244513	19	CEY80D3	Caenorhabditis elegans	3.75e-04

ALIGNMENTS

RESULT 1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS	Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
DEFINITION	Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
ACCESSION	L29052				
VERSION	9456416				
KEYWORDS	L29052.1 GI:456416				
SOURCE	cell death protein; interleukin-1 beta converting enzyme.				
ORGANISM	Caenorhabditis elegans (strain N2) DNA.				
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.				
AUTHORS	Yuan, J., Shaham, S., Jedoux, S., Ellis, H.M. and Horvitz, R.				
TITLE	The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1b-converting enzyme				
JOURNAL	Cell 75, 641-652 (1993)				
MEDLINE	94061982				
FEATURES	Location/Qualifiers				
SOURCE	1..7653				
	/organism="Caenorhabditis elegans"				
	/strain="N2"				
	/db_xref="taxon:6239"				
	2167..2366				
	/gene="ced-3"				
	/number=1				
gene	join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075)				
	/gene="ced-3"				
	join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075)				
CDS	/gene="ced-3"				
	/note="codes for a protein similar to mammalian interleukin-1b-converting enzyme"				
	/codon_start=1				
	/product="cell death protein"				

US-08-287-669-18-04.rst

RESULT	15	D75308	360 bp	mRNA	EST	14-DEC-1995
LOCUS		CEK00884F	Yuji Kohara unpublished	CDNA	Caenorhabditis elegans	CDNA
DEFINITION		clone yk984 5', mRNA sequence.				
ACCESSION		D75308				
NTD		G1121092				
VERSION		D75308.1	GI:1121092			
KEYWORDS		EST.				
SOURCE		Caenorhabditis elegans.				
ORGANISM		Caenorhabditis elegans.				
		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;				
REFERENCE		1 (bases 1 to 360)				
AUTHORS		Kohara,Y., Mitsukuri,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabata,H.				
TITLE		Toward an expression map of the C.elegans genome				
JOURNAL		Unpublished (1994)				
COMMENT						

FEATURES	source
sequence: stop: 363.	
location/Qualifiers	
1..360	
/organism="Caenorhabditis elegans"	
/strain="CB1489 him-8(e1489)"	
/note="dev-stage=varied, sex=Hermaphrodite male	
tissue_type=whole animal"	
/db_xref="taxon:6239"	
/clone="YK89h4"	
/clone_lib="Yuji Kohara unpublished cDNA"	
107 a 81 c 52 g 119 t 1 others	
BASE COUNT	
ORIGIN	

	Query Match	3.88;	Score 38;	DB 32;	Length 360;
	Best Local	Similarity 83.1%;	Fed. No. 4.20e-13;		
	Matches	54; Conservative	0; Mismatches 10;	Indels 1;	Gaps 1
Db	82	TACGTCGTCATTAAAGATTACGTAGTATTTCTACACGAAATATTTGGACGTCAAC	141		
Cp	3214	TACATATGTCATTTTAAAGGCTACTGTATTTTTCGTACGAGATATTTCCG-CGTCAAT	3156		
Db	142	ATGTT	146		
Cp	3155	ATGTT	3151		

Search completed: Fri Aug 6 05:15:58 1999
Job time : 1741 secs.

Db	1	TTTGCGCTAAGAGATCTTGGCCGTAAGTATGGGCAATACGACATTCACAGATTTT	60
Cp	3184	TTTGCGCTACGAGAT-TTTGGCCGCTCAATATGTTGAAGTAATGCCATTCACAGATTTA	3126
Db	61	ATGTT	65
Cp	3125	GTGTT	3121
RESULT	12		
LOCUS	C29898	300 bp	mRNA
DEFINITION	C29898 Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA		EST
ACCESSION	C29898		
NID	92361694		
VERSION	C29898.1	GI:2361694	
KEYWORDS	EST.		
SOURCE	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans.		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.		
AUTHORS	1 (bases 1 to 300) Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.		
TITLE	Expression map of the C.elegans genome		
JOURNAL	Unpublished (1996)		
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1393504.		
FEATURES			
source	Source		
	Contact: Yui Kohara Gene Library Lab National Institute of Genetics Yata 111, Mishima, Shizuoka 411, Japan Tel: 0559-75-0721 Fax: 0559-75-6240 Email: ykohara@dbj.nig.ac.jp. Location/Qualifiers		
	1..300		
	/organism="Caenorhabditis elegans"		
	/strain="CB1489 him-8(e1489)"		
	/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"		
	/db_xref="taxon:6239"		
	/clone="YK228a6"		
	/clone_lib="Yui Kohara unpublished CDNA"		
	111 a 42 c 60 g 86 t 1 others		
BASE COUNT			
ORIGIN			
Query Match	3.8%;	Score 38;	DB 15; Length 300;
Best Local Similarity	90.4%;	Pred. No. 4.20e-13;	
Matches	47;	Conservative 0;	Mismatches 4; Indels 1; Gaps 1;
Db	164	AAATGACATCTGTAATTTTCGCTACGAGATATTTTGGCGTCAATATGTTG	215
Cp	3200	AAAGGCTACTGTAATTTTCGCTACGAGATATT-GCGGCTCAAAATATGTTG	3150
RESULT	13		
LOCUS	C32017	300 bp	mRNA
DEFINITION	C32017 Yui Kohara unpublished cDNA Caenorhabditis elegans CDNA		EST
ACCESSION	C32017		
NID	92363813		
VERSION	C32017.1	GI:2363813	
KEYWORDS	EST.		
SOURCE	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans.		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis		
AUTHORS	1 (bases 1 to 300) Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.		
TITLE	Expression map of the C.elegans genome		

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 252)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eumphylophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

1 (bases 1 to 252)
Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, D.I., Kang, K.Y., Lee, M.C. and Eun, M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797457.

CONTACT: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.astei.re.kr
Submitted by Baek Hie Naim, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bnamhdbioserver.myongji.ac.kr
Seq primer: M13 Reverse primer.
Location/Qualifiers

FEATURES
SOURCE
1.252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: Bluescript SK(+); Site_1: EcoRI, Site_2: XhoI; Directional CDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT
ORIGIN
5 a 21 c 12 g 35 t 179 others

Query Match
Best Local Similarity 10.7%; Pred. No. 8.12e-29;
Matches 25; Conservative 117; Mismatches 88; Indels 3; Gaps 3;
DB 17 AAMTSTYCHGNBVMVCAVSHGNYKSVNCTBRGHCCKRVNNSYTMWGVMBNBSGD 76
CP 3310 AAAATATCAAAAGAAAGAAACAAAATTAATTAATTCGTCACAAATTCGAGCAT 3251
DB 77 WHYMEVNTKVDVGNHTRCSMRBVTYRAHYHDYTCBYYNNNDYHHMBBMYBB-TGCM 135
CP 3250 AATTCGTTTCGAATAATGACCCCTAATGACACTACATAGCATTTAAAGGTTAC 3191
DB 136 TCTMCMCBHYNTR-CTASGMHTSTNYDVKSTN-TWGVTSYDKSMAGYCSBVKYHTK 193
CP 3190 TGTATTTTCCTACAGATATTTTCGCGCTCAAAATATGTAGTAATCGCATCTCAGA 3131
DB 194 VSTTATSTSYCVKRYCYMMMTKKVKKYHYVBBCBHTDSCKTMMNTNKHV 246
CP 3130 ATTAGTGTGCATCAACATCAAGGTTGCTGGGAATGCTGACTGGAG 3078

RESULT 5
LOCUS C41501 357 bp mRNA EST 10-SEP-1997
DEFINITION C41501 Yuj1 Kohara unpublished CDNA Caenorhabditis elegans CDNA
ACCESSION C41501
NID 92377738
VERSION C41501.1 GI:2377738
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 357)
AUTHORS Kohara, Y., Motobashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,

TITLE
JOURNAL
COMMENT

1 (bases 1 to 357)
Sano, M., Miyata, A. and Nishigaki, A.
Expression map of the C.elegans genome Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1402115.

CONTACT: Yuj1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbi.nig.ac.jp.
Location/Qualifiers

FEATURES
SOURCE
1.357
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="YK268f11"
/clone_lib="Yuj1 Kohara unpublished CDNA"
BASE COUNT
ORIGIN
95 a 90 c 49 g 123 t

Query Match
Best Local Similarity 89.5%; Pred. No. 3.19e-20;
Matches 51; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
DB 15 ATCATGCTACAGTACATGCTATTAAGGTTTCTAGTATTTGCTGCGAGATATTG 71
CP 3222 ATGCAGCTACATAGTATTTAAAGGTTTCTAGTATTTTTCGTCACGAGATATTG 3166

RESULT 6
LOCUS T01235 359 bp mRNA EST 10-NOV-1992
DEFINITION WEST01956 Early embryo, Strata gene (cat. #937007) Caenorhabditis elegans CDNA clone CEES041, mRNA sequence.
ACCESSION T01235
NID 9277716
VERSION T01235.1 GI:277716
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 359)
McCombie, W.R., Kelley, J.M., Aubin, L., Goscoechea, M., Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R., Venter, J.C. and Fields, C.A.
Caenorhabditis elegans CDNAs Unpublished (1993)
Other ESTs: WEST01957
Contact: Kerlavage, AR Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 30186990423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.
Location/Qualifiers

FEATURES
SOURCE
1.359
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="CEES041"
/clone_lib="Early embryo, Strata gene (cat. #937007)"
BASE COUNT
ORIGIN
115 a 61 c 60 g 122 t 1 others

Query Match
Best Local Similarity 81.4%; Pred. No. 3.19e-20;
Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch.un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 04:46:57 1999; Maspar time 1724.51 Seconds
1358.744 Million cell updates/sec

Tabular output not generated.

Title: >US-08-287-669-18
(2701-3700) from US08287669.seq (4 of 10)
Description: 1000
Perfect Score: 2701 TTGACGTGAATATATTTA.....GTAAGGAGCATGATTT 3700
N.A. Sequence: 2701 AACCTGACCTTATATTAAT.....CATTTTGGGATGATGAA
Comp: AACCTGACCTTATATTAAT.....CATTTTGGGATGATGAA

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_gss1 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.614; Variance 3.619; scale 3.209

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
C 1	63	6.3	359	8 T01235	WEST01956 Early embryo	4.74e-40
C 2	62	6.2	377	15 C51090	C51090 Yuij Kohara unp	6.60e-39
C 3	53	5.3	253	17 AA754459	97SN1787 Rice Immature	8.12e-29
C 4	53	5.3	253	17 AA754459	97SN1787 Rice Immature	8.12e-29
C 5	45	4.5	359	8 T01235	WEST01956 Early embryo	3.19e-20
C 6	45	4.5	359	8 T01235	WEST01956 Early embryo	3.19e-20
C 7	44	4.4	357	15 C41501	C41501 Yuij Kohara unp	3.52e-19
C 8	43	4.3	247	17 AA754458	97SN1784 Rice Immature	3.81e-18
C 9	43	4.3	247	17 AA754458	97SN1784 Rice Immature	3.81e-18
C 10	38	3.8	300	15 C52036	C52036 Yuij Kohara unp	4.20e-13

C 11	38	3.8	300	15 C53842	C53842 Yuij Kohara unp	4.20e-13
C 12	38	3.8	300	15 C29898	C29898 Yuij Kohara unp	4.20e-13
C 13	38	3.8	300	15 C32017	C32017 Yuij Kohara unp	4.20e-13
C 14	38	3.8	337	32 D64660	CELK086DZR Yuij Kohara	4.20e-13
C 15	38	3.8	360	32 D75308	CELK098H4F Yuij Kohara	4.20e-13
C 16	38	3.8	2275	20 AF034173	AF034173 Human mRNA (T	4.20e-13
C 17	38	3.8	2275	20 AF034173	AF034173 Human mRNA (T	4.20e-13
C 18	38	3.8	2275	20 AF034173	AF034173 Human mRNA (T	4.20e-13
C 19	37	3.7	300	15 C29898	C29898 Yuij Kohara unp	4.00e-12
C 20	37	3.7	337	32 D64660	CELK086DZR Yuij Kohara	4.00e-12
C 21	37	3.7	360	32 D75308	CELK098H4F Yuij Kohara	2.06e-07
C 22	32	3.2	300	15 C32017	C32017 Yuij Kohara unp	2.06e-07
C 23	32	3.2	300	15 C32017	C32017 Yuij Kohara unp	2.06e-07
C 24	31	3.1	807	21 C29734	CELK098H4F Yuij Kohara	1.64e-06
C 25	30	3.0	300	15 C29734	CELK098H4F Yuij Kohara	1.26e-05
C 26	30	3.0	408	22 AD070726	AD070726 Yuij Kohara unp	1.26e-05
C 27	29	2.9	448	37 AD003622	AD003622 Yuij Kohara unp	1.26e-05
C 28	29	2.9	255	22 AI059267	AI059267 Yuij Kohara unp	9.30e-05
C 29	29	2.9	300	15 C35709	C35709 Yuij Kohara unp	9.30e-05
C 30	29	2.9	401	15 C52036	C52036 Yuij Kohara unp	9.30e-05
C 31	29	2.9	420	24 AA740614	AA740614 Yuij Kohara unp	9.30e-05
C 32	29	2.9	429	17 AA740614	AA740614 Yuij Kohara unp	9.30e-05
C 33	29	2.9	559	26 AU005020	AU005020 Yuij Kohara unp	6.59e-04
C 34	28	2.8	163	30 C45014	C45014 Yuij Kohara unp	6.59e-04
C 35	28	2.8	252	21 C94106	C94106 Dictyostellium d	6.59e-04
C 36	28	2.8	430	21 AA971574	AA971574 Yuij Kohara unp	6.59e-04
C 37	28	2.8	562	26 AU001283	AU001283 Dictyostellium d	6.59e-04
C 38	28	2.8	620	25 C22780	C22780 Dictyostellium d	6.59e-04
C 39	28	2.8	854	20 C89844	C89844 Dictyostellium d	4.47e-03
C 40	27	2.7	164	20 C89844	C89844 Dictyostellium d	4.47e-03
C 41	27	2.7	300	15 C35137	C35137 Yuij Kohara unp	4.47e-03
C 42	27	2.7	300	9 AA216925	AA216925 Yuij Kohara unp	4.47e-03
C 43	27	2.7	371	12 AA390241	AA390241 Yuij Kohara unp	4.47e-03
C 44	27	2.7	410	26 AI394542	AI394542 Yuij Kohara unp	4.47e-03
C 45	27	2.7	574	41 AQ347729	AQ347729 Yuij Kohara unp	4.47e-03

ALIGNMENTS

RESULT	1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															</
--------	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

FT FT /*tag= p
FT FT /rpl_type= INVERTED
FT FT /note="Repeat 3"
FT exon 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT mutation 5757
FT FT /*tag= r
FT FT /note="G>A, from allele n2433"
FT FT /*tag= s
FT FT /number= Intron_5
FT exon 5815..5942
FT FT /*tag= t
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= u
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= v
FT FT /rpl_type= INVERTED
FT exon 6298..6537
FT FT /*tag= w
FT FT /number= Exon_7
FT intron 6538..7012
FT FT /*tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= y
FT FT /rpl_type= INVERTED
FT FT /note="Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= z
FT FT /rpl_type= INVERTED
FT FT /note="Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= aa
FT FT /number= Exon_8
PN W09325685-A.
PD 23-DEC-1993.
PE 14-JUN-1993. U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-879638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR P-PDB; R53281.
PT Isolated c elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryonic but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 2701 ttgaagtgaaatataatttatttacttgaagctcgagtgattattttttacacta 2760
Qy 2701 TTGAAGTGAATAATATTTATTATTACTGAAGCTCGAGTGATTATTTTAACTA 2760
Db 2761 atttcgtgagcgaagagccattttagatttgcgaaataacttgcacacacac 2820
Qy 2761 ATTTCGTGAGCGCAAAAGCCATTTTTGTAGATTGCGGAAATACCTGTACACACAC 2820
Db 2821 aacacacatctcttcaataatcccttttccagtggtgacgcaatgctgcaattcga 2880
Qy 2821 ACACACATCTCTTCATCAATATCCCTTTTCCAGTGTTGATCGAATCTGTCAATTTCCA 2880
Db 2881 gtgtccaatgtcaccggcgaagccatgctggagccgcgcaattgaagcccgcgctacac 2940
Qy 2881 GTGTCCAAATGTCAACGGGCAAGCCATCGTCGAGCCGCGCATTTGAGCCCGCGGTAAAC 2940
Db 2941 ttcaaccgaccgagttcacggtgacagcgctcttcaagtgatcaatcaattcttata 3000
Qy 2941 TTCACGACCGCGAGTTTACCGTGACACCGCTCTTCACTGTCATCACTTCTTATCA 3000
Db 3001 ggatactactcaagagcaagatcgttctcgatcgctgacatcattcatcgatcgatcg 3060
Qy 3001 GGATATCTACTCAAGACAGCAAGATCTGCTTCGATCGCGGTGACCTTCATTCATCGGATG 3060
Db 3061 acaaatattatctctctccatcaagcaagatttccacgcaaccttgaatgttgcg 3120
Qy 3061 ACACATATTTATCTCTCTCTCAAGTCAAGCATTTCCACGCCAACCCTTGTATGTGATGCG 3120
Db 3121 aacactaaatctcgaagaatgcgcatctacacataattgagcgcaaatatctctgacg 3180
Qy 3121 AACACTAAATTTTGAGATGCGCATTTACTACATATTTTGACCGCAATATCTCGTACG 3180
Db 3181 gaaaataacagtaacccttaataatgactatgtagtcgattacgagctcgatttcg 3240
Qy 3181 GAAATAATACGATACCCCTTAAATGACTATTTGATGATGATGATGATGATGATGATG 3240
Db 3241 aaacgaataatgctcgaatttgcgaacgaatttcaatttgcatttgcgttctt 3300
Qy 3241 AAACGATATATCTCTGAAATGTGCAACGAATTTAAATTTGCAATTTTGTGTTTCT 3300
Db 3301 ttgataatttgcataatcaataatatttccgtgaacagacacacgagctgacgact 3360
Qy 3301 TTGATATTTTGTGATCAATTAATAATATTTCGTAAGACACACGCGCTACGTAAT 3360
Db 3361 ctttaagaagttacagtagtttgcgttcaagataatttgaagaatlttaaacattt 3420
Qy 3361 CTTTAAAGATTTACAGTGTGTTTGTGCTTCAAGATTTTGAAGAATTTTAAACATTT 3420
Db 3421 tgaaaaaaatacatctcaaatgtgcgaacgcttttcaagtttcgacgatttttga 3480
Qy 3421 TGAATAAAATATATCTACATATGCTCCAAACCTTTTTCAGATTTCCAGATTTTGA 3480
Db 3481 ctttttcaatcaagatagcttattaaacataatcaatcaatcaatcaatcaatcaat 3540
Qy 3481 TTTTTTCAATCAAGATATGCTTATTAAACATATATATATATATATATATATATAT 3540
Db 3541 tagaaatttgcgttcttgcgttcttgcgttcttgcgttcttgcgttcttgcgttctt 3600
Qy 3541 TAGAAATTTTGGCTTTTGTGCTTGTGATGCTTCAATTTTGAATTTGCTCAACAAAAA 3600
Db 3601 tcaatgttcttgcataatgaagcaaaatagcaaatlttataatlttccctat 3660
Qy 3601 TCATGTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3660
Db 3661 tcaatgttgcagaataatgaataaagcgatgattt 3700
Qy 3661 TCATGTGTGCGAGAAATAGTAAAAAGCCGATGATTT 3700

Query Match	99.6%	Score 996	DB 9	Length 7653
Best Local Similarity	99.8%	Pred. No.	0.00e+00	
Matches 998	Conservative	0	Mismatches 2	Indels 0
				Gaps 0

[illegible]

RESULT	ID	13	standard; DNA; 7653 BP.
AC	DT	064737;	standard; DNA; 7653 BP.
DE	DT	23-JUN-1994	(first entry)
KW	DT	ced-3 (G5757A)	gene.
KW	DT	C. elegans; ced-4; ced-3; mutant;	transcriptional regulation
KW	DT	hypogonensis; cell death; hydrophilic;	transmembrane; region
KW	DT	hydrophobic; mutation; amino acid;	substitution; RNA splicing
OS	DT	protein synthesis; null phenotype;	calcium-binding domain; ss
OS	DT	Caenorhabditis elegans.	
FH	DT	Key	Location/Qualifiers
FT	DT	repeat_unit	1356..1472
FT	DT	/*tag= a	/rpt_type= INVERTED
FT	DT	/note= "Inverted w.r.t repeat at 1490-1614"	
FT	DT	1490..1614	
FT	DT	/*tag= b	/rpt_type= INVERTED
FT	DT	/note= "Inverted w.r.t. repeat at 1356-1472"	
FT	DT	2232..2366	
FT	DT	/*tag= c	/number= Exon_1
FT	DT	2367..2429	
FT	DT	/*tag= d	/number= Intron_1
FT	DT	2450..2575	
FT	DT	/*tag= e	/number= Exon_2
FT	DT	2576..2853	
FT	DT	/*tag= f	/number= Intron_2
FT	DT	2854..3107	
FT	DT	/*tag= g	/number= Exon_3
FT	DT	3108..4302	
FT	DT	/*tag= h	/number= Intron_3
FT	DT	3126..3243	
FT	DT	/*tag= i	/rpt_type= INVERTED
FT	DT	/note= "Inverted w.r.t. repeat at 3329-3396"	
FT	DT	3329..3396	
FT	DT	/*tag= j	/rpt_type= INVERTED
FT	DT	/note= "Inverted w.r.t. repeat at 3126-3243"	
FT	DT	3487..3759	
FT	DT	/*tag= k	/rpt_type= INVERTED
FT	DT	/note= "Inverted w.r.t. repeat at 3782-4070"	
FT	DT	3782..4070	
FT	DT	/*tag= l	/rpt_type= INVERTED
FT	DT	/note= "Inverted w.r.t. repeat 3487-3759"	
FT	DT	4303..4634	
FT	DT	/*tag= m	/number= Exon_4
FT	DT	4635..5546	
FT	DT	/*tag= n	/number= Intron_4
FT	DT	4688..4719	
FT	DT	/*tag= o	/rpt_type= INVERTED
FT	DT	/note= "repeat 3"	
FT	DT	5221..5330	
FT	DT	repeat_unit	

```

FT FT      /*tag= a
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 1490-1614"
FT FT      repeat_unit
FT FT      /*tag= b
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT      exon
FT FT      /*tag= c
FT FT      /number= Exon_1
FT FT      mutation
FT FT      /*tag= d
FT FT      /note= "C>T, from allele n1040"
FT FT      intron
FT FT      /*tag= e
FT FT      /number= Intron_1
FT FT      exon
FT FT      /*tag= f
FT FT      /number= Exon_2
FT FT      mutation
FT FT      /*tag= g
FT FT      /note= "G>A, from allele n718"
FT FT      intron
FT FT      /*tag= h
FT FT      /number= Intron_2
FT FT      exon
FT FT      /*tag= i
FT FT      /number= Exon_3
FT FT      intron
FT FT      /*tag= j
FT FT      /number= Intron_3
FT FT      repeat_unit
FT FT      /*tag= k
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT      repeat_unit
FT FT      /*tag= l
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT      repeat_unit
FT FT      /*tag= m
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT      repeat_unit
FT FT      /*tag= n
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT FT      repeat_unit
FT FT      /*tag= o
FT FT      /number= Exon_4
FT FT      intron
FT FT      /*tag= p
FT FT      /number= Intron_4
FT FT      repeat_unit
FT FT      /*tag= q
FT FT      /rpt_type= INVERTED
FT FT      /note= "Repeat 3"
FT FT      repeat_unit
FT FT      /*tag= r
FT FT      /rpt_type= INVERTED
FT FT      /note= "Repeat 3"
FT FT      exon
FT FT      /*tag= s
FT FT      /number= Exon_5
FT FT      mutation
FT FT      /*tag= t
FT FT      /note= "G>A, from allele n2433"
FT FT      intron
FT FT      /*tag= u
FT FT      /number= Intron_5
FT FT      exon
FT FT      /*tag= v
FT FT      /number= Exon_6

```

```

FT FT      mutation
FT FT      /*tag= w
FT FT      /note= "C>T, from allele n1165"
FT FT      intron
FT FT      /*tag= x
FT FT      /number= Intron_6
FT FT      repeat_region
FT FT      /*tag= y
FT FT      /rpt_type= INVERTED
FT FT      mutation
FT FT      /*tag= z
FT FT      /note= "G>A, from allele n717"
FT FT      exon
FT FT      /*tag= aa
FT FT      /number= Exon_7
FT FT      mutation
FT FT      /*tag= ab
FT FT      /note= "C>T, from allele n1949"
FT FT      mutation
FT FT      /*tag= ac
FT FT      /note= "G>A, from allele n1286"
FT FT      mutation
FT FT      /*tag= ad
FT FT      /note= "C>T, from allele n1129 and n1164"
FT FT      mutation
FT FT      /*tag= ae
FT FT      /note= "C>T, from allele n2430"
FT FT      mutation
FT FT      /*tag= af
FT FT      /note= "G>A, from allele n2426"
FT FT      intron
FT FT      /*tag= ag
FT FT      /number= Intron_7
FT FT      repeat_unit
FT FT      /*tag= ah
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT      repeat_unit
FT FT      /*tag= ai
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT FT      exon
FT FT      /*tag= aj
FT FT      /number= Exon_8
FT FT      mutation
FT FT      /*tag= ak
FT FT      /note= "C>T, from allele n1163"
FT FT
FT FT      WO9325685-A.
FT FT      PD 23-DEC-1993.
FT FT      PF 14-JUN-1993; U05701.
FT FT      PR 12-JUN-1992; US-897788.
FT FT      PR 20-NOV-1992; US-979638.
FT FT      PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT FT      PI Horvitz HR, Shahan S, Yuan J;
FT FT      DR WPI; 94-007542/01.
FT FT      DR P-PSDB; RA7466.
FT FT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT      develop agents to increase or prevent cell death in organisms
FT FT      PS Clam 2; Fig 4; 127bp; English.
FT FT      CC This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
FT FT      CC was identified as the ced-3 transcript and was most abundant in
FT FT      CC embryos, but was also detected in larvae and young adults, suggesting
FT FT      CC that ced-3 is expressed not only in cells undergoing cell death. The
FT FT      CC four largest introns as well as sequences 5' of the start codon
FT FT      CC contain repetitive elements, some of which have been characterised
FT FT      CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
FT FT      CC and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
FT FT      CC highly hydrophilic with no significant hydrophobic region that might
FT FT      CC be a transmembrane region. One region of Ced-3 is very rich in serine.
FT FT      CC It is thought that this region is involved in protein-protein inter-
FT FT      CC actions, similar to acid blobs in transcription factors. Of the
FT FT      CC mutations which occur within the ced-3 gene, eight of the mutations
FT FT      CC are missense mutations, two are nonsense mutations and two are putative

```

FT intron 5943..6297
 FT /tag= t
 FT /number= Intron_6
 FT repeat_region 6062..6138
 FT /tag= u
 FT /rpt_type= INVERTED
 FT 6298..6537
 FT /tag= v
 FT /number= Exon_7
 FT mutation 6322
 FT /tag= w
 FT /note= "C>T, from allele n1949"
 FT intron 6538..7012
 FT /tag= x
 FT /number= Intron_7
 FT repeat_unit 6567..6625
 FT /tag= y
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT repeat_unit 6905..6965
 FT /tag= z
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT exon 7012..7075
 FT /tag= aa
 FT /number= Exon_8
 FT
 PN W09325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR P-PSDB; R53283.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 develop agents to increase or prevent cell death in organisms
 PS Claim 14; Fig 4; 127pp; English.
 CC The sequences given in Q64735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterized in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC functions is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterized by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, petrogenic
 CC infection, aging or hair loss.
 SO Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2821 acacacatctcctcaaatccctttccagtgctgactcgaatgctgacatcga 2880
 Qy 2821 AACACATCTCTTCAAAATATCCCTTTTCCAGTGTGACTCGAATGCTGCAATTGCA 2880
 Db 2881 gttccaatgtacacgagcaagccatcgtcgaagccgcatattgagcccgccgctaacac 2940
 Qy 2881 GTTCCAAATGTCACGGCAAGCCATGTCGAGCCGCGCATATGAGCCCCGCGGTACAC 2940
 Db 2941 ttacccgaccgagttcacccgagcagcgtctcttcagtgatcatattcattcttaca 3000
 Qy 2941 TTCACCGACCGAGTTCACCGTACAGCGCTCTTCAAGTGTATATTCCTTTTACA 3000
 Db 3001 ggtatctactcaagaagcaagatctcgtctcgtacgcgtgcatcttaccagtcagc 3060
 Qy 3001 GGATATCTACTCAAGCAAGATCTGCTCGATGCGGTGCACTTCAATTCGATCGATCG 3060
 Db 3061 acacaattatcattctctccagtcgaagcattccagccaaacctgtatgtatgagc 3120
 Qy 3061 ACACAATTATTCATCTCTCCAGTCAGTCMAAGCATTTCCACCACTTGATGTATGCG 3120
 Db 3121 aacactaaattctgagaatgagcattactcaacatatattgacgagcaaatctcgtac 3180
 Qy 3121 AACACTAAATTCGAGAAATGCGATTACTCAACATTTTACGCGCAATATCTGTAAC 3180
 Db 3181 gaaaatacagtaaccctttaaatactatgtagtgctgattacagggctcgatttcg 3240
 Qy 3181 GAAAATACAGTAAACCTTTAAATGACTATTGAGTGTGATTAAGGCTCGATTTTCG 3240
 Db 3241 aaacgaatatatgctcgaattgtgacacgaatttcaattgtcatttttggtttctt 3300
 Qy 3241 AAACGAATATATCTCGAATTGTGACACGAATTTTAAATTTTGCATTTTGTGTTTCT 3300
 Db 3301 ttgatattttgataaataaataaataattcttcgtaacagacacacgagtcgact 3360
 Qy 3301 TTGATATTTTGATCAATATATATATATTTTCGTAACAGACACGCGTACAGTACT 3360
 Db 3361 cttaaagaagttacagtagtttctcgttcaagatatattgaaaagaattttaaactt 3420
 Qy 3361 CTTTAAAGAGTTACAGTATGTTTCCCTCAAGATTTTGAAGAAATTTAAACATTT 3420
 Db 3421 tgaaaaaaatacttaacatgagccaaagcgttttccaagtttcgagatttttga 3480
 Qy 3421 TGAATAAAATTCATCACTAACATGTCGCAAAACGCTTTTTCATTTTCGAGATTTTGA 3480
 Db 3481 ttttttcattcaagatgattcatttaacacatatattcaacttaattgaaattcttg 3540
 Qy 3481 TTTTTCATTCAAGATATCTTATTAACACATATATATATCTTATGATGATTTCTTG 3540
 Db 3541 tagaaatttgggctttcgtctcgtatgctcactttgaaattgctcaagcaaaaaa 3600
 Qy 3541 TAGAAATTTTGGGCTTTCTGCTCTAGATGCTCTACTTTGAATATGCTCAACGAAAAA 3600
 Db 3601 tcatgtggttgcataatgataagcaaaaaaataagaatttttatattttccctat 3660
 Qy 3601 TCATGTGTTGTTGTTCAATATGACGAAAAAATGCAATTTTATATATTTTCCCTAT 3660
 Db 3661 tcatgtggtgcagaaaaatagtaaaaaagcgatgatt 3700
 Qy 3661 TCATGTTGTGCAAAAAATAGTAAAAAAGCGCATGCAATT 3700

RESULT 12
 ID 054666 strand: DNA; 7653 BP.
 AC 054666: 23-JUN-1994 (first entry)
 DT 23-JUN-1994 (first entry)
 DE ced-3 gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT repeat_unit 1356..1472


```
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT repeat_region 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT mutation 6434
FT /tag= w
FT /note= "C>T, from allele n1129 and n1164"
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
PN MO9325685-A.
PD 23-DEC-1993.
PE 14-JUN-1993; 005701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-879638.
PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shiham S, Yuan J;
PI WPI: 94-007542/01.
DR P-PDB: R53285.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PS develop agents to increase or prevent cell death in organisms
PS Claim 14: Fig 4: 127pp: English.
CC The sequences given in 064735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as tem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
```

```
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 bp: 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaaatattatttattactgaagtcgagtgatatttattttaaacta 2760
Qy 2701 TTGAAGTGAATATATTATTTTATTACTGAAGCTCGAGTGATTTATTTTAACTA 2760
Db 2761 atttcgtgacgcaaaagccatttctgtagatttgcgaaataacttgcacacacac 2820
Qy 2761 ATTTCTGGCGCAAAAGCCATTGTGTAGATTGGCGAAATATTGTCACACACACAC 2820
Db 2821 acaacatcttcctcaaatatcccttttccagtggttgactggaatgctgcaattcga 2880
Qy 2821 ACACACATCTCTTCAAAATATCCCTTTTCCAGTGTGACTCGAAATGCTGCAATTGCA 2880
Db 2881 gtgtccaatgtcaacggcaagccatcgtcgagcgcgcataggcccgcggtacac 2940
Qy 2881 GTGTCCAATGTCAACGGCAAGCCATCGTGGAGCCCGCATTTAGCCCGCGCTACAC 2940
Db 2941 ttcacgacccgagatcccggtacagcgctctctagtgatcatcttacttctatca 3000
Qy 2941 TTCACGACCCGAGATTACCGGTACAGCGCTCTTCAAGTCAATCACTTCACTTATCA 3000
Db 3001 ggatatctactcaaggaagaatcctggttctgtagtcgctgacatcatcatcgagatcg 3060
Qy 3001 GGATATCTACTCAAGACCAAGATCTGTTCTGATGCGCGGCACTTCAATCACTCGGATCG 3060
Db 3061 acacaattatcatctccctccagtcacaagcattcccaagcaaccttgatgttgatcg 3120
Qy 3061 ACACAATTATTCATCTCTCTCCAGTCAAGCATTTCCACGCAACCTTGATGTGATCGG 3120
Db 3121 aacactaaattctgagaatgagcatctactcaacataattgacgcgcaaatatctcgtacg 3180
Qy 3121 AACACTAAATTCTGAGATGCGCATTAACAATTTTGACGCGCAAAATATCTGTACG 3180
Db 3181 gaaaaataacagtaacctttaaagtactatgagtgctgagattacgggctcgatttcg 3240
Qy 3181 GAAAAATACAGTAACCTTTAAATGACTATTTGATGTGATTTAGGGCTCATATTTGCG 3240
Db 3241 aaacgaatatatctcgaattgagcaagcaatttaatttgcatttttggttttctt 3300
Qy 3241 AAACGATATATGCTCGAATTGGACACGAATTTTAATTTTGCATTTTGTGTTTCTT 3300
Db 3301 ttgatatttttgatcaattataaattattccgttaaacagacacagcgatcagttact 3360
Qy 3301 TTGATATTTTTCATCAATTAATAATTAATTTTCGTTAAACAGACACAGCGGTACAGTACT 3360
Db 3361 cttttaaagtgtaacgtagtcttgccttcaagatatcttgaagaagaattttaaactt 3420
Qy 3361 CTTTAAAGAGTTACGATTAATTTTCCTTCAGATATTTTGAAAAAGAAATTTTAAACATTT 3420
Db 3421 tgaaaaaaatactcaacatgagcgaagcgttttccaaagtttcgaagtttttga 3480
Qy 3421 TGAATAAAATTCATCTACATGTGCCAAAAGCGTTTTCAGATTTTCGAGATTTTGTGA 3480
Db 3481 ttttttcaatcaagatalgctattataacacatatataatcatcttaagtgtgaattcttg 3540
Qy 3481 TTTTTCATTCAAAGATATGCTATTAATTAACACATATTAATATCATTAATGTAATTTCTTG 3540
Db 3541 tagaagaatttgggcttcttgcttctagatgcttacttcttgaattgctcaagaataaa 3600
Qy 3541 TAGAATTTTGGCTTCTTCTTCTAGATGCTCTACTTTTGAAATTTGCAACGAAATAAA 3600
Db 3601 tcatgtggttgcctatataagaccaaaataagcaatttttatatttccctct 3660
Qy 3601 TCATGTGTTTGTTCATATGAAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
```

CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterized by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaatatatttttactgaaagctgagatatttttttaacacta 2760
 QY 2701 TTGAAGTGAATATATTTTACTGAAAGCTGAGATATTTTAACTTAACACTA 2760
 Db 2761 atttcgfgggcaaaagccattttagatttgcgaaatacttgcacacacac 2820
 QY 2761 ATTTGCTGGCCAAAGGCCATTTTGATTTGCCGAAATCTTGTACACACACAC 2820
 Db 2821 aacacatccctccaatatcccttttccagttgtactcgaaatgttcgaatcga 2880
 QY 2821 AACACATCTCTTCAATATCCCTTTTCCAGTGTGACTGCAATGCTGCGAATTCGA 2880
 Db 2881 gtgtccaatgtccggcaagccatcgttcgagacgagcaatgagcccgccgagctaac 2940
 QY 2881 GTGTCCAATGTCAACGGCAAGCCATGTCGAGCCGCGCATTTGACCCCGCGCTACAC 2940
 Db 2941 ttccagcccgagttcccgagagagtccttcagttgtatattatcattcattatca 3000
 QY 2941 TTCCAGCCGAGTTCCCGAGAGGCTCTTCCAGTGTATATTTACTCTTATCA 3000
 Db 3001 ggaatactcaagaagaagctcgttcgttcgacgcgctgacttcatcgagctg 3060
 QY 3001 GGATATCTATCAAGAGCAAGATCTGTCGTGATCGCGTGCATCTTCAATCGAGATCG 3060
 Db 3061 acacaattatcatctctccagtcagcaagcatttcccgacaacctgtatgtatcg 3120
 QY 3061 ACACAAATTTATCATCTCTCCAGTCAACGCAATTTCCAGCCAACTTGATGATCG 3120
 Db 3121 aacactaaatttggagaagcgcatctacacatatattgagcgcaaatatctctgtagc 3180
 QY 3121 AACACTAAATTTGAGAAATGCGCATTTACTCAACATATTTGAGCGCAATATCTGTAGC 3180
 Db 3181 gaaataacagtaaccctttaaatgactatgtatgttcgaatttcagggtcgcgatttcg 3240
 QY 3181 GAAATAATACGTAACCTTTAATGACTATTTAGTGTGATTTACGGGCTCGATTTTCG 3240
 Db 3241 aaacgaatatatgctcgaatgtgacaacgaatttaatttgcatttctgtgttctt 3300
 QY 3241 AAACGAATATATGCTCGAATTTGACAACGAATTTAATTTGTCAATTTTGTGTTCTT 3300
 Db 3301 ttgatatttttgcataataataattttccgtataaagaacacagcggtacagctact 3360
 QY 3301 TTGATATTTTATCATATTAATTAATTTATTTCCGTAAGACACACGCGTACGTAAT 3360
 Db 3361 cttttaaagatttacagttgttcgcttcacagatatatttgaagaagaatttaaacattt 3420
 QY 3361 CTTTAAAGATTACAGTGTTCCTTCACAGATATTTTGAAGAAATTTTAAACATTT 3420
 Db 3421 tgaataaaataatcatcaaatgtgcgaacagcttttcaagtttcgagaatttttga 3480
 QY 3421 TGAATAAAATAATCATCAATGTGCGCAAAACGCTTTTTCAGAGTTTCGAGATTTTGA 3480
 Db 3481 ttttttcatcaagatatgcttattacaacataataataatcaattgaattcttg 3540
 FT

QY 3481 TTTTTCATTCAGATATGCTTATTAACACATATATATCATTAATGCAATTTCTTCG 3540
 Db 3541 tagaaatttgggcttccttcgcttcagtagctcacttgaatgtcacaagaaaaa 3600
 QY 3541 TAGAAATTTGGCTTTTCGCTTCTAGTACTGCTACTTTGAATGTGCACGAAAAA 3600
 Db 3601 tcatgtgttcttcataatgaaatgacccaataaagaatttttataatcttcccat 3660
 QY 3601 TCATGTGTTGTTTCATATGAAATGACGAAAAATGCAATTTTATATATTTTCCCTAT 3660
 Db 3661 tcatgtgttcagaataatagtaaaagcgcatcatt 3700
 QY 3661 TCATGTTGTCCAAAAAATAGTAAAAACGCGCATGCAATTT 3700

RESULT 9
 ID 064742 standard; DNA; 7653 BP.
 AC 064742;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (C6434T) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 PH Key
 FT repeat_unit
 FT 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1490-1614"
 FT 1490..1614
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /tag= e
 FT /number= Exon_2
 FT 2576..2853
 FT /tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= m
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= n


```

FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT      repeat_unit
FT      /tag= j
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT      repeat_unit
FT      /tag= k
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT      repeat_unit
FT      /tag= l
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT      exon
FT      /tag= m
FT      /number= Exon_4
FT      /tag= n
FT      /number= Exon_4
FT      /tag= o
FT      /number= Intron_4
FT      repeat_unit
FT      /tag= o
FT      /rpl_type= INVERTED
FT      /note= "Repeat 3"
FT      repeat_unit
FT      /tag= p
FT      /rpl_type= INVERTED
FT      /note= "Repeat 3"
FT      exon
FT      /tag= q
FT      /number= Exon_5
FT      /tag= r
FT      /number= Intron_5
FT      /tag= s
FT      /number= Exon_6
FT      /tag= t
FT      /number= Intron_6
FT      mutation
FT      /tag= u
FT      /note= "C>T, from allele n717"
FT      repeat_region
FT      /tag= v
FT      /rpl_type= INVERTED
FT      /tag= w
FT      /number= Exon_7
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit
FT      /tag= y
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon
FT      /tag= aa
FT      /number= Exon_8

```

```

PS      Claim 14: Fig 4: 127pp: English.
CC      The sequences given in Q64735-45 represent mutations of the ced-3 gene.
CC      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
CC      Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
SQ
Query Match          99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-NOV-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI: 94-007542/01.
 DR P-PSDB: R53287.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 14; Fig 4; 127pp; English.
 CC The sequences given in 064735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 CC Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaataatatttacttgaagtcgagtgattatttttaacta 2760
 QY 2701 ttgaagtgaataatatttacttgaagtcgagtgattatttttaacta 2760
 Db 2761 atttggtggcgaagggcatttggatgttgcgaaatacttccacacac 2820
 QY 2761 atttggtggcgaagggcatttggatgttgcgaaatacttccacacac 2820
 Db 2821 acacacatctctcaaatatcccttccagtggtgactgcatgctgcaatcga 2880
 QY 2821 acacacatctctcaaatatcccttccagtggtgactgcatgctgcaatcga 2880
 Db 2881 gttgccaatgtcaccggaagcattgctgcgacgacgacgacgacgacgac 2940
 QY 2881 gttgccaatgtcaccggaagcattgctgcgacgacgacgacgacgacgac 2940
 Db 2941 ttccacgaccggaagtccacgacgacgacgacgacgacgacgacgacgac 3000
 QY 2941 ttccacgaccggaagtccacgacgacgacgacgacgacgacgacgacgac 3000
 Db 3001 ggatattctcaagaagaatctgcttcgacgacgacgacgacgacgacgac 3060
 QY 3001 ggatattctcaagaagaatctgcttcgacgacgacgacgacgacgacgac 3060
 Db 3061 acacaaatttcaatctctccagtcacgacgacgacgacgacgacgacgac 3120
 QY 3061 acacaaatttcaatctctccagtcacgacgacgacgacgacgacgacgac 3120
 Db 3121 aaacaaatttcaagaagaatctgacgacgacgacgacgacgacgacgacgac 3180
 QY 3121 aaacaaatttcaagaagaatctgacgacgacgacgacgacgacgacgacgac 3180
 Db 3181 gaataatacgaatccttaataatgactatgtagtgcattacgacgacgacgac 3240
 QY 3181 gaataatacgaatccttaataatgactatgtagtgcattacgacgacgacgac 3240

Db 3241 aaacgaatatatgctcgaattgacacgaatatttaatttgcatttctgttctc 3300
 QY 3241 aaacgaatatatgctcgaattgacacgaatatttaatttgcatttctgttctc 3300
 Db 3301 ttgatattttgacaaataataatatttccgtaaacagcaccagcgtacgact 3360
 QY 3301 ttgatattttgacaaataataatatttccgtaaacagcaccagcgtacgact 3360
 Db 3361 ctttaaaagatcagatgatttgccttcaagatatatttgaagaatatttaacatt 3420
 QY 3361 ctttttaaaagatcagatgatttgccttcaagatatatttgaagaatatttaacatt 3420
 Db 3421 tgaataaaatcactcaatcagatgacacaaagccttttccaagttcgcagatttga 3480
 QY 3421 tgaataaaatcactcaatcagatgacacaaagccttttccaagttcgcagatttga 3480
 Db 3481 ttttttcaatcagatgatttgccttcaatcactaataatatttgaatttctg 3540
 QY 3481 ttttttcaatcagatgatttgccttcaatcactaataatatttgaatttctg 3540
 Db 3541 tagaattttggccttcgcttcgacgacgacgacgacgacgacgacgacgacgac 3600
 QY 3541 tagaattttggccttcgcttcgacgacgacgacgacgacgacgacgacgacgac 3600
 Db 3601 tcatgtgttgcacatgacacgacacacacacacacacacacacacacacacac 3660
 QY 3601 tcatgtgttgcacatgacacgacacacacacacacacacacacacacacacac 3660
 Db 3661 tcatgtgttgcac 3700
 QY 3661 tcatgtgttgcac 3700

RESULT 7
 ID 064739 standard; DNA: 7653 BP.
 AC 064739:
 DT 23-JUN-1994 (first entry)
 DE ced-3 (66297A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT repeat_unit 1356..1472
 FT /*tag= a
 FT /rpt_type= INVERTED
 FT /note="inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /*tag= b
 FT /rpt_type= INVERTED
 FT /note="inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /*tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /*tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /*tag= e
 FT /number= Exon_2
 FT 2576..2853
 FT /*tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /*tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /*tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /*tag= i

QY 3061 ACACATATATCATCTCCCTCCAGTCAACGCATTTCCAGCCAACTGTATGTGATGCG 3120
Db 3121 aacctaatcttagaatgctgcttactcaactatttgaacgagcaatatctcgtagc 3180
W 3121 AACCTAAATCTGAGATGCGCATTTACTCAACATATTTGACGGGAAATCTCGTAGC 3180
Db 3181 gaaataacagtaaccccttaaaagactatgttagtgcgaattcgaattcgcgaattcgc 3240
QY 3181 GAAATAATACATACCTTAATGACTATGTAGTGTGATTTAGCGGCTCGATTTTCG 3240
Db 3241 aaacgaatatactcgaattgtgacaaacgaatttcaattgttcaatttgcgttctc 3300
QY 3241 AAACGAATATATGCTCGATTTGACAAACGAATTTATTTGATTTGTTGTTCTT 3300
Db 3301 ttgatattttgatcaataataattattcgttaaacgacccagcggtacagtagc 3360
QY 3301 TTGATATTTTTCATCAATTAATTAATTTCCGTAACACAGACACGCGCTACAGTACT 3360
Db 3361 ctttaagaagctacagtagtttcgcttcaagatatttgaagaatttcaacattc 3420
QY 3361 CTTTAAAGAGTTCAGAGTTCCTTCGCTCAAGATATTTGAAAAGATTTTAAACATT 3420
Db 3421 tgaataaataatcatcacaatgycacaaacgcttttcaagttcgcagatttttga 3480
QY 3421 TGAATAAATAATCATCTAAACATGTGCCAAACGCTTTTTCAGATTTCGAGATTTTGA 3480
Db 3481 ttttttcatccaagatatgcttataacacataattatcatatgtgaatttctg 3540
QY 3481 TTTTTCATTCACAGATATGCTTTATTAACATTAATTTATCATTAATGTGAATTTCTG 3540
Db 3541 tagaattttgagcttttcgctcagtagctcactcttgaatttgaatttgaacaaata 3600
QY 3541 TAGAATTTTGGGCTTTTCGTTAGTATGCTCTACTTTTGAATTTGCTCAACGAAAAA 3600
Db 3601 tcatgtggttgcataatgaatgacaaataatgacaatttataatttccctat 3660
QY 3601 TCATGTGTTTGTTCATATGATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgtgtgcagaaaaatagtaaaaaagcgcatgcttt 3700
QY 3661 TCATGTGTGTCAGAAAAATAGTAAAAAAGCGCATGATTT 3700

RESULT 6
ID 064744 standard; DNA; 7653 BP.
AC 064744;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6536A) gene.
KW C. elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2332..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /*tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /*tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /*tag= f
FT FT Intron

FT FT /number= Intron_2
FT FT 2854..3107
FT FT /*tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /*tag= h
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /*tag= i
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /*tag= j
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /*tag= k
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /*tag= l
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /*tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /*tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /*tag= o
FT FT /rpl_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /*tag= p
FT FT /rpl_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /*tag= r
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= s
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= t
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= u
FT FT /rpl_type= INVERTED
FT FT 6298..6537
FT FT /*tag= v
FT FT /number= Exon_7
FT FT 6535
FT FT /*tag= w
FT FT /note= "G>A, from allele n2426"
FT FT 6538..7012
FT FT /*tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= y
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= z
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= aa
FT FT /number= Exon_8
FT FT W09325685-A.

```

FT FT /number= Exon_1
FT FT 2367..2429
FT FT /tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /tag= f
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /tag= h
FT FT /number= Intron_3
FT FT 3126..3343
FT FT /tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /tag= q
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /tag= r
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /tag= s
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /tag= t
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /tag= u
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /tag= v
FT FT /number= Exon_7
FT FT 6372
FT FT /tag= w
FT FT /note= "G>A, from allele n1286"
FT FT 6538..7012
FT FT /tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /tag= y
FT FT /rpt_type= INVERTED

```

```

FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /tag= z
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /tag= aa
FT FT /number= Exon_8
FT FT W09325685-A.
FT FT 23-DEC-1993.
FT FT 14-JUN-1993; U05701.
FT FT 12-JUN-1992; US-897788.
FT FT 20-NOV-1992; US-979638.
FT FT (MAST ) MASSACHUSETTS INST TECHNOLOGY.
FT FT Horvitz HR, Shaham S, Yuan J;
FT FT WPI; 94-007542/01.
FT FT P-PDB; R53284.
FT FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT develop agents to increase or prevent cell death in organisms
FT FT Claim 14; 127pp; English.
FT FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT FT was most abundant in embryos, but was also detected in larvae and young
FT FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT FT cell death. The four largest introns as well as sequences 5' of the
FT FT start codon contain repetitive elements, some of which have been
FT FT characterised in non-coding regions of other C. elegans genes, such
FT FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT FT region that might be a transmembrane region. One region of Ced-3 is
FT FT very rich in serine. It is thought that this region is involved in
FT FT protein-protein interactions, similar to acid blobs in transcription
FT FT factors. Of the mutations which occur within the ced-3 gene, eight of
FT FT the mutations are missense mutations, two are nonsense mutations and
FT FT two are putative splicing mutations. These mutations establish the
FT FT null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT FT function is not essential for viability. The ced-3 and ced-4 gene
FT FT products may be used to develop agents for treating conditions
FT FT characterised by cell deaths, such as myocardial infarction, stroke,
FT FT CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT FT CC infection, aging or hair loss.
FT FT Sequence 7653 BP: 2430 A; 1453 C; 1271 G; 2499 T;
FT FT
FT FT Query Match 99.6%; Score 996; DB 9; Length 7653;
FT FT Best Local Similarity 99.8%; Pred. No. 0.00e+00;
FT FT Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

FT      exon      6298..6537
FT      /tag= w
FT      /number= Exon_7
FT      intron     6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit 6567..6625
FT      /tag= y
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit 6905..6965
FT      /tag= z
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon       7012..7075
FT      /tag= aa
FT      /number= Exon_8
FT
FT      .W09325685-A.
FT      23-DEC-1993.
FT      14-JUN-1993; U05701.
FT      12-JUN-1992; US-897788.
FT      20-NOV-1992; US-979638.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Horvitz HR, Shahan S, Yuan J;
FT      WPI: 94-007542/01.
FT      P-RSDB: R53282.
FT
FT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT      develop agents to increase or prevent cell death in organisms
FT      PS Claim 14: Fig 4; 127pp; English.
FT      The sequences given in 064735-45 represent mutations of the C. elegans
FT      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT      was most abundant in embryos, but was also detected in larvae and young
FT      adults, suggesting that ced-3 is expressed not only in cells undergoing
FT      cell death. The four largest introns as well as sequences 5' of the
FT      start codon contain repetitive elements, some of which have been
FT      CC characterised in non-coding regions of other C. elegans genes, such
FT      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT      CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT      CC region that might be a transmembrane region. One region of Ced-3 is
FT      CC very rich in serine. It is thought that this region is involved in
FT      CC protein-protein interactions, similar to acid blobs in transcription
FT      CC factors. Of the mutations which occur within the ced-3 gene, eight of
FT      CC the mutations are missense mutations, two are nonsense mutations and
FT      CC two are putative splicing mutations. These mutations establish the
FT      CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT      CC function is not essential for viability. The ced-3 and ced-4 gene
FT      CC products may be used to develop agents for treating conditions
FT      CC characterised by cell deaths, such as myocardial infarction, stroke,
FT      CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT      CC infection, aging or hair loss.
FT      S2 Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

```

```

Query Match      99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2941 TTACCGACCCGAGTTCCACCGTACAGCGCTCTTCACTGATATCACTTCTTATCA 3000
DB      3001 ggaatcactcaagaagaagatcctcgtctcgtatcgctgacatcattcattcgcgacg 3060
QY      3001 GGATATCTACTCAAGAGCAAGATCTGTTCTCGATGGCTGCATTCATTCATTCGGATCG 3060
DB      3061 acaaatattcactcctccacgtcaagaagcattccacgaacttctgtatgtgacg 3120
QY      3061 ACACAATTATTCATCTCCCTCCAGTCAACCAATTTCCACCAACCTGTATTTGATGCG 3120
DB      3121 aaacaaattcctgagatgagcattaccacacatatattgagcgcaaatcctcgtgac 3180
QY      3121 AACACTAAATTCGAGAAATGCCATTACTCAACATTTTGAACGCCCAATATCTGTAAC 3180
DB      3181 gaaaaatacagtaacccttaataatgactatgtagtgctgacttaacggctcgtatcgc 3240
QY      3181 GAAAAATACAGTACCCCTTTAATGACTATGTAGTGTGATTTACGGGCTCGATTTTCG 3240
DB      3241 aaacgaatatagctcgaattgtgacacgaatttaatttcaatttctgtgttctc 3300
QY      3241 AAACGAATATATGCTCGAATGTGACAAACGAATTTTAATTTGTCAATTTTGTGTTTCTT 3300
DB      3301 ttgatatatttgaatcaataataataattatcttcgtaaacgacacacgagcgtaagctact 3360
QY      3301 TTGATATTTTGTGACATTAATTAATTAATTTATTCGTAACAGACACGCGCTACAGTACT 3360
DB      3361 ctcttaaaagatcacagtagttcttcgctcaagaatatttgaanaaatcttaacattc 3420
QY      3361 CTTTAAAGAGTTACAGTAGTGTGCTTCAAGATATTTTGAAGAATTTTAAACATTT 3420
DB      3421 tgaanaaaatcactaactgacatgacaaagccttttcaagttcgcagatcttttga 3480
QY      3421 TGAANAATAATCATCTAACATGTCGCAAAACCTTTTTCAGTTTCGACAGATTTTGA 3480
DB      3481 ttttttcttaagatgactatataacacatatataatttaagaatgagattcttg 3540
QY      3481 TTTTTCATTCAGATATGACTTATTTAAACATATTAATTAATTAATGATGTAATTCG 3540
DB      3541 tagaaatttgggcttcctcgtcctcgtacgtcctactttgaattgctcaacgaanaaa 3600
QY      3541 TAAAAATTTGGGCTTTTGGTTCTGTAGTACTCTACTTTTGAATTTGCTCAACGAAAAA 3600
DB      3601 tcatgtgttgcctcatalgaaatgacaaataatagcaattttatataatttccctcat 3660
QY      3601 TCATGTGTTTGTGTCATATGATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
DB      3661 tcatgtgtgcgaanaaaatagtaaaaaagcgatcattt 3700
QY      3661 TCATGTGTGCAAAAAATATGTAATAAAGCGATGCATTT 3700

```

RESULT 5

ID 064741 standard; DNA: 7653 BP.

```

AC      064741;
DT      23-JUN-1994 (first entry)
DE      ced-3 (G6372A) gene.
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain; ss.
OS      Caenorhabditis elegans.
FH      key
FH      repeat_unit 1356..1472
FT      /tag= a
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 1490-1614"
FT      repeat_unit 1490..1614
FT      /tag= b
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 1356-1472"
FT      exon      2232..2266
FT      /tag= c

```


Db	2821	aacacatctcttcoaatatcccccttttcgaagtgtagctcgaaatgctgtgcgaattcga	2880
Oy	2821	ACACACATCTCCTTAAATAATCCCTTTTCAGTGTTACTCGAATTGCTGTGAATTTTGA	2880
Db	2881	gtgtccaatgtcacccgcaagccalcgcgcgagccgcgcgatitgaagccccgcgcgtacac	2940
Oy	2881	GTTGTCAATGTACACGGGAGAAGCATGTGTGGAGCCGCCAATTGAGCCCCGGGCTACAC	2940
Db	2941	ttaaccggaccggagttaacgctgtgacaagcgtctctcagtggtcatcatcaacttcta	3000
Oy	2941	TTCACCGACCAGTGCACCGTAGACAGGCTCTTCAATGTCAATCATTCACCTTCATCA	3000
Db	3001	ggatactacctcaagaagaagatcctcgtctccgatgcgcgtgcacttatcatcgatcg	3060
Oy	3001	GGATTATCTACTCACAGACAGACATCTCGTTTCGATGCGGCTGCACCTTCATCATCGATCG	3060
Db	3061	acacaaattatcatctcctccagtlcaacgcataltcccagccaacctgtatgtatgcg	3120
Oy	3061	ACACAATATTCATCTCCTCCAGTCACGACATTTMCCAGCAACCTGTATGTGATGCG	3120
Db	3121	aacctaaattctagaagatgcgcattactcaaacattataggcgcgcaatatctcgtagc	3180
Oy	3121	AACACTAAATCTGAGATGCGCATTTACTCAACATATTGAGCGCCAATATCTGTAGC	3180
Db	3181	gaaaaatacaglaaacctttaatgatcatttgtagtgltgcgaattiaagcgctcgatllcg	3240
Oy	3181	GAAAATATCAGTAACCCCTTTAAATGACTATTGTAGTGTGATTTACGGGCTCGATTTTCG	3240
Db	3241	aaaagaaataatgctcgcgaattgtgcacaacgaatttaatttgtcaatttttgtttctc	3300
Oy	3241	AAACGAATAATMGCTCGAATTTGTGACAAGAAATTTTAATTTTGTCATTTTGTGTTTTCTT	3300
Db	3301	tgtatatttctgatacaataataataatttcocgrtaaacagacacagcgcgtacagta	3360
Oy	3301	TGTATATTTTGTGATCAATATTAAATTAATTTCOCGRAAACAGACACGCTACAGTACT	3360
Db	3361	cttttaaagagttacagtagtttcgcgtccaagataatttgaaaagaattttaaacatt	3420
Oy	3361	CTTTTAAAGAGTTACGATGTTTGTGCTCAAGAAATTTTGAAAAGAATTTTAAACATTT	3420
Db	3421	tgaaaaaaacatcatcaatcagtgccaaacgccttttccaagttgcagatllttga	3480
Oy	3421	TGAAAAAAATCATCTACATCAATGTGCCAAACGCTTTTTCAGTTTCCACAGATTTTGTGA	3480
Db	3481	tttttttcatcaagatagtcttaliaacacataataatlatcoataytgaattcttg	3540
Oy	3481	TTTTTTTCATTCAGATATGCTTATTATAACATATTAATTAATCATTAATGTGAATTTCTTG	3540
Db	3541	tagaaatttgagccttcggttcttagtgtctctacttligaatttgtlcaaagaaaaa	3600
Oy	3541	TAGAAATTTTGGGCTTTTGTCTAGTATGCTTACCTTTTGAAATTTGTCAACGAAAAAA	3600
Db	3601	tcaigtggttgcgtcataatgacgccaaaaaataagcaatttltalacatlltccccat	3660
Oy	3601	TCANGTGTTGTTCATATGATGACGAAAAATACCAATTTTTTNATATTTTCCCCTAT	3660
Db	3661	tcatgtgtgcgaagaaaaatagtaaaaaagcgcatgcatllt 3700	
Oy	3661	TCATGTTGTGCAAGAAAAATAGTAAAAAGCCGATCCATTTT 3700	
FT	/*tag= a		
FT	/rpl_type= INVERTED		
FT	/note= "Inverted w.r.t repeat at 1490-1614"		
FT	repeat_unit		
FT	1490..1614		
FT	/*tag= b		
FT	/rpl_type= INVERTED		
FT	/note= "Inverted w.r.t. repeat at 1356-1472"		
FT	2232..2366		
FT	/*tag= c		
FT	/number= Exon_1		
FT	2367..2429		
FT	/*tag= d		
FT	/number= Intron_1		
FT	2450..2575		
FT	/*tag= e		
FT	/number= Exon_2		
FT	2576..2853		
FT	/*tag= f		
FT	/number= Intron_2		
FT	2854..3107		
FT	/*tag= g		
FT	/number= Exon_3		
FT	3108..4302		
FT	/*tag= h		
FT	/number= Intron_3		
FT	3126..3243		
FT	/*tag= i		
FT	/rpl_type= INVERTED		
FT	/note= "Inverted w.r.t. repeat at 3329-3396"		
FT	3329..3396		
FT	/*tag= j		
FT	/rpl_type= INVERTED		
FT	/note= "Inverted w.r.t. repeat at 3126-3243"		
FT	3487..3759		
FT	/*tag= k		
FT	/rpl_type= INVERTED		
FT	/note= "Inverted w.r.t. repeat at 3782-4070"		
FT	3782..4070		
FT	/*tag= l		
FT	/rpl_type= INVERTED		
FT	/note= "Inverted w.r.t. repeat 3487-3759"		
FT	4303..4634		
FT	/*tag= m		
FT	/number= Exon_4		
FT	4635..5546		

RESULT	4	
ID	064738	standard; DNA: 7653 BP.
AC	064738;	
DT	23-JUN-1994	(first entry)
DE	ced-3 (C5940T) gene.	
KW	<i>C. elegans</i> ; ced-4; ced-3; mutant; transcriptional regulation;	
KW	emryogenesis; cell death; hydrophilic; transmembrane; region;	
KW	hydrophobic; mutation; amino acid; substitution; RNA splicing;	
KW	protein synthesis; null phenotype; calcium-binding domain; ss.	
OS	<i>Caenorhabditis elegans</i> .	
FT	Key	Location/Qualifiers
FT	repeat-unit	1356..1472
FT	exon	5815..5942
FT	/*tag=	s
FT	/number=	Exon_6
FT	5757	
FT	/*tag=	t
FT	/note=	"C>T, from allele n1165
FT	5943; 6297	
FT	/*tag=	u
FT	/number=	Intron_6
FT	repeat_region	6062..6138
FT	/*tag=	v
FT	/rpt_type=	INVERTED

NT 23-JUN-1994 (first entry)
 DE ced-3 (G2487A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key
 FT repeat_unit location/Qualifiers
 FT 1356..1472
 FT /tag= a
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /tag= b
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /tag= e
 FT /number= Exon_2
 FT 2487
 FT /tag= f
 FT /note= "G>A, from allele n718"
 FT 2576..2853
 FT /tag= g
 FT /number= Intron_2
 FT 2854..3107
 FT /tag= h
 FT /number= Exon_3
 FT 3108..44302
 FT /tag= i
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= j
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= k
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= l
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= m
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= n
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= o
 FT /number= Intron_4
 FT 4688..44719
 FT /tag= p
 FT /rpl_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= q
 FT /rpl_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /tag= r
 FT /number= Exon_5
 FT 5761..5814
 FT /tag= s
 FT /number= Intron_5

FT exon 5815..5942
 FT /tag= t
 FT /number= Exon_6
 FT 5943..6297
 FT /tag= u
 FT /number= Intron_6
 FT 6062..6138
 FT /tag= v
 FT /rpl_type= INVERTED
 FT 6298..6537
 FT /tag= w
 FT /number= Exon_7
 FT 6538..7012
 FT /tag= x
 FT /number= Intron_7
 FT 6567..6625
 FT /tag= y
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /tag= z
 FT /rpl_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7012..7075
 FT /tag= aa
 FT /number= Exon_8
 FT
 FT W09325685.A.
 FT PD 23-DEC-1993.
 FT PF 14-JUN-1993; U05701.
 FT PR 12-JUN-1992; US-897788.
 FT PR 20-NOV-1992; US-979638.
 FT PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 FT PI Horvitz HR, Shaham S, Yuan J;
 FT DR WPI: 94-007542/01.
 FT DR P-PSDB: R53280.
 FT PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 FT develop agents to increase or prevent cell death in organisms
 FT PS Claim 14; Fig 4; 127pp; English.
 FT CC The sequences given in G64735-45 represent mutations of the C. elegans
 FT CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 FT CC was most abundant in embryos, but was also detected in larvae and young
 FT CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 FT CC cell death. The four largest introns as well as sequences 5' of the
 FT CC start codon contain repetitive elements, some of which have been
 FT CC characterised in non-coding regions of other C. elegans genes, such
 FT CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 FT CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 FT CC region that might be a transmembrane region. One region of Ced-3 is
 FT CC very rich in serine. It is thought that this region is involved in
 FT CC protein-protein interactions, similar to acid blobs in transcription
 FT CC factors. Of the mutations which occur within the ced-3 gene, eight of
 FT CC the mutations are missense mutations, two are nonsense mutations and
 FT CC two are plative splicing mutations. These mutations establish the
 FT CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 FT CC function is not essential for viability. The ced-3 and ced-4 gene
 FT CC products may be used to develop agents for treating conditions
 FT CC characterised by cell deaths, such as myocardial infarction, stroke,
 FT CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 FT CC infection, aging or hair loss.
 FT SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
 Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 2701 ttgaagtgaaataatattactgaagctgagatgtattttttaaaccta 2760
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Oy 2701 TTGAAGTGAATAATATTATTACTGAAGCTGAGATGATATTATTTTAACTA 2760
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 2761 atttcgtgcgcaaaagccattttagattgcgaaataactgtcacacacac 2820
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Oy 2761 ATTTCGTGCGCAAAAGCCATTTTTGTAGATTTCGCGAAATACCTGTGCACACACAC 2820

```

FT      Intron      5943..6297      /*tag= x
FT      /label= Intron-6
FT      repeat_unit 6062..6138      /*tag= y
FT      /label= Repeat-4
FT      exon        6298..6537      /*tag= z
FT      allele      6322
FT      /tag= aa
FT      /label= T(n1949)
FT      /note= "Creates premature stop codon at 412"
FT      allele      6372
FT      /tag= ab
FT      /label= A(n1286)
FT      allele      6434
FT      /tag= ac
FT      /label= T(n1129,n1164)
FT      /note= "causes A449V mutation"
FT      allele      6485
FT      /tag= ad
FT      /label= T(n2430)
FT      /note= "causes A466V mutation"
FT      allele      6535
FT      /tag= ae
FT      /label= A(n2426)
FT      /note= "causes E483K mutation"
FT      intron      6538..7012
FT      /tag= af
FT      /label= Intron-7
FT      repeat_unit 6567..6625
FT      /tag= ag
FT      /label= Repeat-5
FT      repeat_unit 6905..6965
FT      /tag= ah
FT      /label= Repeat-5
FT      exon        7013..7653
FT      /tag= ai
FT      /note= "in-frame stop codon at 7073-7075"
FT      allele      7020
FT      /tag= aj
FT      /label= T(n1163)
FT      /note= "causes S486F mutation"
FT      allele      709625946-A1
FT      PN          23-AUG-1996
FT      PF          23-FEB-1996: U02473
FT      PR          24-FEB-1995: US394189
FT      PA          (MASI) MASSACHUSETTS INST TECHNOLOGY
FT      PI          Horvitz HR, Shaham S, Yuan J;
FT      DR          WPI: 96-425082/42
FT      P-PSDB: R98754
FT      PT          Ced-3 and human interleukin 1-beta convertase genes and proteins
FT      useful to treat inflammation and diseases characterised by cell
FT      death
PS      Claim 18: Fig 3; 139pp; English.
CC      The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
CC      mapping RFLPs and chromosome walking, and genomic DNA cloned
CC      in plasmid pJ107 was sequenced. EMS+-induced alleles were also
CC      sequenced. The gene codes for a cell death protein (R98754) that
CC      is structurally similar to human interleukin-1 beta converting
CC      enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
CC      cysteine protease like ICE and that ICE may be a human equivalent
CC      of the nematode cell death gene. The ced-3 gene can be used as a
CC      probe or in the prodn. of Ced-3 protein and novel drugs for
CC      enhancing or inhibiting the activity of ICE, ced-3 and related
CC      genes for the treatment of inflammatory diseases and/or diseases
CC      caused by cell death. Novel inhibitors of ced-3 activity include
CC      portions of the ced-3 gene and its product.
SQ      Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;

```

Query Match 100.0%; Score 1000; DB 24; Length 7653;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db      2701 ttgaagtgaaatatattttattacgaaagctcgagtgattatttttaacacta 2760
      |||
      2701 TTGAAGTGAATATATATTTTATTACTGAAAGCTCGAGTGATTTATTTTAACTA 2760
Db      2761 atttcgtggcgaaagagcattttagatttgcgaaataacttgcacacacac 2820
      |||
      2761 ATTTTCGTGGCGGCAAAAGCCATTTTGTAGATTTCGGAATAACTTGTACACACACAC 2820
Db      2821 acacacatctcttcaaatatcccttllccagtggttgaactcgaatgcgtgcgaattcga 2880
      |||
      2821 ACACACATCTCTCTCAAAATATCCCTTTTTCAGATGTGACTGCAATGCTGCAATTCGA 2880
Db      2881 gtttcacaaatgcacggcgaagccatcgtcgagcgcgcaatlgagcccgcggtacac 2940
      |||
      2881 GTTTCCAATGTCAACCGGCAAGCCATCGTCGAGCCCGCATTTAGCCCGCGGTACAC 2940
Db      2941 ttccacgaccggagttcacgctgacagcgctctcttgaagtgatcaattcaattctatca 3000
      |||
      2941 TTCACGACCCGAGTTTACCGCTACAGCGCTCTTTCAGTGTCAATTCATCTTATCA 3000
Db      3001 ggaatctactcaagagcaagatctcgttcgctgacgcgtgcacatcattcactgcgacgc 3060
      |||
      3001 GGATATCTACTCAAGAGCAAGATCTCGTCTCGATCGCGGTGCACTTCATTCATGCGATCG 3060
Db      3061 acacaaattatctatctccctccagtaacagcattcccagccaacctgtatgttgatgcg 3120
      |||
      3061 ACACAAATTATTCATCTCCCTCACTCAACGATTTCCAGCCAACTTGTATGTGATGCG 3120
Db      3121 aacacataattcgaagaatcgagattacacataatttgaacggcaaatatctcgtacg 3180
      |||
      3121 AACACATAATTCTGAGAAATCGGATTTACTACACATTTTGGCCGCAATATCTGATACG 3180
Db      3181 gaaataatcacgtaacccctttaaigtactttagtgatgatttgaagcggtcgatlltcg 3240
      |||
      3181 GAAATAATCACGTAAACCTTTAAATGACTATTGATGATGATTTAGCGGCTGATTTTGG 3240
Db      3241 aaacgaataatagctcgaatgtgacacgaatattcaattgtcatlltgggttctt 3300
      |||
      3241 AAACGAATATATCTCGAAATGTGACACGAATTTAAATTTGTCATTTTGTGTTTCTT 3300
Db      3301 ttgataatttgaatcaataataattatcttcgtaaaagacacacagcgctacagttact 3360
      |||
      3301 TTGATATTTTGTGATCAATTATATTAATTTTCCGTAAACAGACACCGCGCTACGTTACT 3360
Db      3361 cttttaaagagttacagtagtttgcgttcaagatalltgaagaagaattttaaacttt 3420
      |||
      3361 CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATTTTGAAGAAGATTTTAAACATTT 3420
Db      3421 tgaataaaataatcatcaaatgacatgycacaaaagccttttccaagtttcgagatlltga 3480
      |||
      3421 TGAATAAAATTCATCTAACATGTGCCAAAACGCTTTTTCAGGTTTCGAGATTTTTCGA 3480
Db      3481 ttttttcatcagaatagctcttaataacacataataataattatcatatgtgaattcttg 3540
      |||
      3481 TTTTTCATTCATCAAGATAGCTTTATTAACATATATATATTAATGTGAATTTCTTG 3540
Db      3541 tagaaatttgggcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 3600
      |||
      3541 TAGAAATTTTGGGCTTGTGCTTGTATGATGCTGCTTGTGAATTTGTCACGAAANAA 3600
Db      3601 tcatgtgtgtgtgtgtatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3660
      |||
      3601 TCATGTGTGTGTGTGTATGAATGAGGAAATAGCAATTTTATATATATTTTCCCTAT 3660
Db      3661 tcatgtgtgtgtgtgtatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3700
      |||
      3661 TCATGTGTGTGTGTGTATGAATGAGGAAATAGCAATTTTATATATATTTTCCCTAT 3700

```

RESULT 3
 ID 064736 standard; DNA: 7653 BP.
 AC 064736;

```

|||||
QY 2701 TTGAAGTGAATAATATTTATTACTGAAAGCGAGTGTATTTATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaaggccatttgtagatttgcggaataacttgcacacacac 2820
QY 2761 ATTTCGTGGCGGCAAAAGGCCATTGTGTAGATTGGCGAAATACTTGTGCACACAC 2820
Db 2821 acacacatctctccaataatcccttttccagtggttgaactgcagatgcgtgaattcga 2880
QY 2821 ACACACATCTCTCTCAAAATATCCCTTTTCCAGTGTGACTGCAATGCTGCAATTCGA 2880
Db 2881 gtttccaatgtcacggcgaagccatgctggagccggcgaattgagccggcggtacac 2940
QY 2881 GTGTCCAAATGTCAACGGGCAAGCCATGCTGGAGCCGGCGATTGAGCCCGCGCTACAC 2940
Db 2941 ttaccgaccggagttcacccgtacagcgtctctcagttgcatcattcattcttata 3000
QY 2941 TTCAACGACCCGAGTTACCGGTGACAGCGTCTCTTCAGTGTCAATTCATCTCTTATCA 3000
Db 3001 ggaatctactcaagagcaagatctcgtctcagtcgctgcagctcattcatcgcgagtcg 3060
QY 3001 GGATATCTACTCAAGACAGATCTGCTCTCGATCGCGCTCACTTCATTCATCGATCG 3060
Db 3061 acacaattatcatctcctccagtcacgcgcatctccagccaaaccttgatgtgtatg 3120
QY 3061 ACACAAATTAATTCATCTCTCCAGTCAACGCAATTTCCAGCAACCTTGATGTTGATGCG 3120
Db 3121 aacacaaattctgagagatcgcatctcaacatattgaacgcaaatatctcgtacg 3180
QY 3121 AACCTTAATTTCTGAGATATGCGCATTTCTCAACATATTTCACGCGCAATATCTCGTAGC 3180
Db 3181 gaaaaatacagtaaacctcttaaatgactatgtagtgatgacatlaagggcgatctcg 3240
QY 3181 GAAAAATACAGTAACTCTTAAATGACTATTGTAGTGTGATTTACGGCGCATTTTCG 3240
Db 3241 aaaaagaatatatgctgcaatgtgacacagaaatttgaatttgcatttggtttctt 3300
QY 3241 AAAAGAAATATATGCTCAAAATGTGACACGAATTTTAAATTTGTCATTTTGTGTTTCCT 3300
Db 3301 ttgatactttgatcaataataaatatatttcgtaaacagacacagcgctacagttact 3360
QY 3301 TTGATATTTTGGATCATTAATAATTAATTTCCCTAAACAGACACCGCGCTACAGTACT 3360
Db 3361 cttttaaagagttacagtagtttcgctcaagatatcttgaagaagaattttaaactt 3420
QY 3361 CTTTAAAGAGTTACAGTAGTTTCGCTCAAGATATTGAAAGAAATTTTAAACATTT 3420
Db 3421 tgaaaaaaaatcatcctaacaatgctcaaaagccttttcaagtttcgacagatttttga 3480
QY 3421 TGAATAAAATCATCTACATGTCCTCAAAACGCTTTTTCAGTTTGCAGATTTTGTGA 3480
Db 3481 tttttcatcagaatagcttaataacacataataatcatcattgaatttcttg 3540
QY 3481 TTTTTCATCATGAGATATGCTTATTAACACATTAATTAATTCATTAATGGAATTTCTTG 3540
Db 3541 tagaaatttgggcttctgctcagtagtctcacttctgaatttgcacagaaaaa 3600
QY 3541 TAGAAATTTGGGCTTTTCGTTCTAGATGCTCTACTCTTGAATTTCTCAACGAAAAAA 3600
Db 3601 tcatgtgtgttgcataatgaacgaagaataagcaattttatatatttccctat 3660
QY 3601 TCATGTGTGTGTTCATATGATGACGAATAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgttgcagaataaataagcagatcatt 3700
QY 3661 TCATGTGTGTGCAAAATAGTAAAAAAGCGCATGCAATTT 3700

```

RESULT 2
 ID T38196 standard; DNA; 7653 BP.
 AC T38196;
 DT 17-DEC-1996 (first entry)
 DE Nematode Ced-3 gene.

```

KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KM cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT   /*tag= a
FT   /label= Repeat-1
FT   /tag= b
FT   /label= Repeat-2
FT   /tag= c
FT   /tag=
FT   /codon_start= 2232..2234
FT   2310
FT   /*tag= d
FT   /label= T(n1040)
FT   /note= "causes L27F mutation"
FT   2367..2429
FT   /*tag= e
FT   /label= Intron-1
FT   /tag= f
FT   2487
FT   /*tag= g
FT   /label= A(n718)
FT   /note= "causes G65R mutation"
FT   2576..2853
FT   /*tag= h
FT   /label= Intron-2
FT   2854..3107
FT   /*tag= i
FT   3108..4302
FT   /*tag= j
FT   /label= Intron-3
FT   3126..3243
FT   /*tag= k
FT   /label= Repeat-1
FT   3329..3396
FT   /*tag= l
FT   /label= Repeat-1
FT   3487..3759
FT   /*tag= m
FT   /label= Repeat-2
FT   3782..4070
FT   /*tag= n
FT   /label= Repeat-2
FT   4303..4633
FT   /*tag= o
FT   4634..5546
FT   /*tag= p
FT   /label= Intron-4
FT   4688..4719
FT   /*tag= q
FT   /label= Repeat-3
FT   5221..5330
FT   /*tag= r
FT   /label= Repeat-3
FT   5546..5760
FT   /*tag= s
FT   5757
FT   /*tag= t
FT   /label= A(n2433)
FT   /note= "causes G360S mutation"
FT   5761..5814
FT   /*tag= u
FT   /label= Intron-5
FT   5815..5942
FT   /*tag= v
FT   5940
FT   /*tag= w
FT   /label= T(n1165)
FT   /note= "creates premature stop codon at 403"
FT

```

```

FT      /tag= g
FT      /note= "G>A, fromm allele n718"
FT      2576..2853
FT      /tag= h
FT      /number= Intron_2
FT      2854..3107
FT      /tag= i
FT      /number= Exon_3
FT      3108..4302
FT      /tag= j
FT      /number= Intron_3
FT      3126..3243
FT      /tag= k
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT      3329..3396
FT      /tag= l
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT      3487..3759
FT      /tag= m
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT      3782..4070
FT      /tag= n
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT      4303..4634
FT      /tag= o
FT      /number= Exon_4
FT      4635..5546
FT      /tag= p
FT      /number= Intron_4
FT      4688..4719
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5221..5330
FT      /tag= r
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5347..5760
FT      /tag= s
FT      /number= Exon_5
FT      5757
FT      /tag= t
FT      /note= "G>A, from allele n2433"
FT      5761..5814
FT      /tag= u
FT      /number= Intron_5
FT      5815..5942
FT      /tag= v
FT      /number= Exon_6
FT      5940
FT      /tag= w
FT      /note= "C>T, from allele n1165"
FT      5943..6297
FT      /tag= x
FT      /number= Intron_6
FT      6062..6138
FT      /tag= y
FT      /rpt_type= INVERTED
FT      6297
FT      /tag= z
FT      /note= "G>A, fromm allele n717"
FT      6298..6537
FT      /tag= aa
FT      /number= Exon_7
FT      6322
FT      /tag= ab
FT      /note= "C>T, from allele n1949"
FT      6372
FT      /tag= ac

```

```

FT      /note= "G>A, from allele n1265"
FT      6434
FT      /tag= ad
FT      /note= "C>T, fromm alleles n1129 and n1164"
FT      6485
FT      /tag= ae
FT      /note= "C>T, from allele n2430"
FT      6535
FT      /tag= af
FT      /note= "G>A, from allele n2426"
FT      6538..7012
FT      /tag= ag
FT      /number= Intron_7
FT      6567..6625
FT      /tag= ah
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      /tag= ai
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7013..7075
FT      /tag= aj
FT      /number= Exon_8
FT      7020
FT      /tag= ak
FT      /note= "C>T, from allele n1163"
FT      1..5850
FT      /tag= al
FT      /note= "Claim 2, inhibitor fragment"
FT      1..3020
FT      /tag= am
FT      /note= "Claim 2, inhibiotr fragment"

```

```

PD      WO9325694-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05705.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-984182.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shahan S, Yuan J;
DR      MPI: 94-007551/01.
DR      P-PSDB; R45262.
PT      Agents which affect activity of cell death genes - used to
PT      develop drugs for treating conditions characterised by cell death
PT      or proliferation
PS      Disclosure; Fig 3: 132pp; English.
CC      This sequence encodes the C.elegans cell death gene, ced-3. Fragments
CC      of the amino terminal of the protein encoded by this sequence act as
CC      inhibitors of ced-3. This gene has considerable similarity to human
CC      interleukin-1beta convertase (ICE), which converts pro-interleukin-
CC      1beta to the active cytokine and is involved in inflammatory response
CC      in humans. The similarity between the two sequences suggests that
CC      inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
CC      nematode Ced-3 proteins have an overall amino acid similarity of 28%.
CC      The ced-3 inhibitors may be used for identifying agents which affect
CC      the activity of a gene belonging to the to the ced-3/ICE family or
CC      genes and for diagnosis of diseases characterised by cell death. They
CC      can also be used to develop drugs for treating conditions characterised
CC      by cell deaths such as myocardial infarction, stroke, degenerative
CC      disease, traumatic brain injury, hypoxia, pathogenic infection, or
CC      hair loss, or drugs for reducing the proliferative capacity or size
CC      of a population of cells such as cancerous cells, cells which produce
CC      autoreactive antibodies, infected cells, hair follicle cells or cells
CC      which are critical to the life of a parasite, pest or recombinant
CC      organism. They may also be used in the diagnosis of inflammatory
CC      disease.
SQ      Sequence. 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 100.0%; Score 1000; DB 9; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 2701 ttgaagtgaataattatttactgaagctcgagtgatattatttttaacacta 2760

```

M P E R E F

(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 05:16:20 1999; Maspar time 234.73 Seconds

Tabular output not generated. 912.853 Million cell updates/sec

Title: >US-08-287-669-18
Description: (2701-3700) from US08287669.seq (4 of 10)
Perfect Score: 1000
N.A. Sequence: 2701 TTGAAGTGAATATATTTA.....GTAAAGACGCATCATTT 3700
Comp: AACTTACTTATATTAAT.....CATTTTTCGCGTAA

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 8.989; Variance 6.868; scale 1.309

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1000	100.0	7653	9	Q54401 ced-3 gene.	0.00e+00
2	1000	100.0	7653	24	T38196 Nematode Ced-3 gene.	0.00e+00
3	996	99.6	7653	9	Q64736 ced-3 (G52487A) gene.	0.00e+00
4	996	99.6	7653	9	Q64738 ced-3 (G59407) gene.	0.00e+00
5	996	99.6	7653	9	Q64741 ced-3 (G6372A) gene.	0.00e+00
6	996	99.6	7653	9	Q64744 ced-3 (G6536A) gene.	0.00e+00
7	996	99.6	7653	9	Q64739 ced-3 (G6297A) gene.	0.00e+00
8	996	99.6	7653	9	Q64745 ced-3 (C7020T) gene.	0.00e+00
9	996	99.6	7653	9	Q64742 ced-3 (C6434T) gene.	0.00e+00

10	996	99.6	7653	9	Q64743 ced-3 (C6485T) gene.	0.00e+00
11	996	99.6	7653	9	Q64740 ced-3 (C6322T) gene.	0.00e+00
12	996	99.6	7653	9	Q54666 ced-3 gene.	0.00e+00
13	996	99.6	7653	9	Q64737 ced-3 (G5757A) gene.	0.00e+00
14	254	25.4	2485	14	Q79970 Interleukin-1 beta co	1.46e-119
15	94	9.4	7653	9	Q64738 ced-3 (C5940T) gene.	3.09e-33
16	94	9.4	7653	24	T38196 Nematode Ced-3 gene.	3.09e-33
17	94	9.4	7653	9	Q64736 ced-3 (G2487A) gene.	3.09e-33
18	94	9.4	7653	9	Q64745 ced-3 (C7020T) gene.	3.09e-33
19	94	9.4	7653	9	Q64739 ced-3 (G6297A) gene.	3.09e-33
20	94	9.4	7653	9	Q64742 ced-3 (G6536A) gene.	3.09e-33
21	94	9.4	7653	9	Q64743 ced-3 (C6434T) gene.	3.09e-33
22	94	9.4	7653	9	Q64741 ced-3 (G6372A) gene.	3.09e-33
23	94	9.4	7653	9	Q64737 ced-3 (G5757A) gene.	3.09e-33
24	94	9.4	7653	9	Q54666 ced-3 gene.	3.09e-33
25	94	9.4	7653	9	Q64744 ced-3 (G6536A) gene.	3.09e-33
26	94	9.4	7653	9	Q64740 ced-3 (C6322T) gene.	3.09e-33
27	94	9.4	7653	9	Q54401 Ced-3 gene.	3.09e-33
28	81	8.1	6560	9	Q54629 Genomic region contai	1.35e-26
29	47	4.7	6560	9	Q54629 Genomic region contai	1.02e-09
30	42	4.2	91	9	Q51746 Oligonucleotide probe	2.16e-07
31	40	4.0	204	1	N81164 Base substituted E.co	1.77e-06
32	38	3.8	91	46	V44650 Mammalian DNA replica	1.41e-05
33	38	3.8	91	9	O51746 Oligonucleotide probe	1.41e-05
34	38	3.8	204	1	N81164 Base substituted E.co	1.41e-05
35	31	3.1	91	46	V44650 Mammalian DNA replica	1.41e-05
36	28	2.8	114	12	Q70468 Generic DNA sequence	2.81e-01
37	28	2.8	114	12	Q70467 Generic DNA sequence	2.81e-01
38	28	2.8	114	12	Q70467 Generic DNA sequence	2.81e-01
39	28	2.8	114	12	Q70467 Generic DNA sequence	2.81e-01
40	28	2.8	501	3	N50024 Sequence encoding new	2.81e-01
41	27	2.7	114	12	Q70469 Generic DNA sequence	7.13e-01
42	27	2.7	114	12	Q70465 Sequence encoding new	1.79e-00
43	26	2.6	501	3	N50023 DNA encoding human MP	1.79e-00
44	26	2.6	2703	32	T59405 DNA encoding human TG	1.79e-00
45	26	2.6	2703	36	T69695 DNA encoding human TG	1.79e-00

ALIGNMENTS

RESULT 1
ID Q54401 standard: DNA; 7653 bp.
AC Q54401;
DT 01-JUL-1994 (first entry)
DE ced-3 gene.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW Interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2356
FT /*tag= c
FT /*number= Exon_1
FT mutation 2310
FT /*tag= d
FT /*note= "C>T, from allele n1040"
FT intron 2367..2429
FT /*tag= e
FT /*number= Intron_1
FT exon 2430..2575
FT /*tag= f
FT /*number= Exon_2
FT mutation 2487

```

/db.xref="PID:e1344529"
/db.xref="PID:g3874792"
/db.xref="GI:3874792"
/db.xref="SPRMBL:P90772"
/translation="MDVDDILGLIILATNGSLAIIALPECGGKNAKSAEGESSLS
SKMGKATPDGREGSEAGTSPSGSAAPAPCGTHDPNQTLAGLGNCCYEKKDK
PAAGCGPAPPAANKPGMAATHDPNQTLAGLNNVFGKDGAAPAAGGGGAPAPADQ
NAKATLHDPNQTLAGLNNVFEKDDGGAPAAIPGSPKAPANGAARATHDPNQT
AGLNNVFEKDDGGAGAGAPGKGPLOPMNKKATLDPNQTLLAAGVDYFAGAK
KVGGAGAPGCKPILOPNNGKATMDPNQTLLAAGVDYFAGADKKK"

```

Query Match	1.2%	Score 112	DB 21	Length 39211
Best Local Similarity	80.6%	Pred. No. 4	38-39	
Matches 233	Conservative	0	Mismatches 43	Indels 13
			Gaps	4
Db	8014	GAACACAAATTTCTGAGATGCGTATTTGCCACACATATTGTACGCGGCAAAATATCTGTA	8073	
Qy	3120	GAACTCTAAATTTCTGAGATGCGCATTTCTCAACATATTGTACGCGCAA-TATCTGTA	3178	
Db	8074	GCGAAACTACAGTATTTCTTTAAATG6CTACTGAGCGCTTATTGATTCACG6GCTC	8133	
Qy	3179	GCGAAATATACGATACCCCTTTAATGACATTTGTAGTG-----TCGATTTACGGGCTC	3232	
Db	8134	CATTTTGAATGAATTTCTTTTGCATATGAGACACAGA-----TATTTGTAATTTTAA	8159	
Qy	3233	GATTTTGCAAAGCAATATATGCTCCAAATGTGACACACAAATTTTATTTTGCAATTTTGT	3292	
Db	8150	GTTTTTCTATATTTT--ATCGATATATAATGATTCGGTAAATGACGCAACGCT	8247	
Qy	3293	GTTTCTTTTGATATTTTGTGATCAATTTATATATTTTCCGTAAACAGACACGAGCGCT	3352	
Db	8248	ACAGTAGCACCGCAATGAATTTCTGTAGTTTCCGTAGAGATATTTTG	8296	
Qy	3353	ACAGTACTTTTAAAGATTACAGTATGTTTTCCGCTTAAAGATATTTTG	3401	

Search completed: Fri Aug 6 04:46:36 1999
Job time : 1857 secs.

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. This sequence is the entire insert of clone C36F7. The true right end of clone C41G7 is at 1443 in this sequence. The start of this sequence (1..103) overlaps with the end of sequence 281048. The end of this sequence (39104..39211) overlaps with the start of sequence 283225.

FEATURES

source location/qualifiers
1..39211
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="I"
/clone="C36F7"
complement(325..3022)
/gene="C36F7.1"
complement(join(325..552,1053..1412,2008..2469,3002..3022))
/gene="C36F7.1"
/note="Predicted using GeneFinder. Similarity to Drosophila araucan homeoprotein (TR:EI218676); cDNA EST YK42311.5 comes from this gene; cDNA EST YK29347.5 comes from this gene; cDNA EST YK455a2.5 comes from this gene"
/codon_start=1
/protein_id="CAB02816.1"
/db_xref="PID:e1344525"
/db_xref="PID:g3874786"
/db_xref="GI:3874786"
/db_xref="SWISS-PROT:Q93348"
/translation="MIPIANOVKMEPTTPQTTPAPPTSPADALRGILPPNLLGOFH FHPMAQVLAQAQAAQAQATLAGEPGVPMPIGADLPKPEMLGGPMPMEF SDARFLHPYGLGKIKKRNATREATAPLMDWLSHHRKNPSKADPMAYGTGMLT OVSTFANARRLKKNKMTWSPQNRGCGDDEDDDDDMKRRSSSTISVSGE SLFGKHLASPTSGSGSGSDLTFRKELVSPFNDSSESPKRRKMSIADVSD DSSKSDSPSSDNLTLTSLFQNAKRMKEVMARLDPTLFRHTMFPFQLAQAQ GQQLALAPDTPTTPNEAGEVSR"
10377..10809
/gene="C36F7.2"
join(10377..10466,10518..10584,10715..10809)
/gene="C36F7.2"
/note="Predicted using GeneFinder"
/codon_start=1
/protein_id="CAB02814.1"
/db_xref="PID:e1344523"
/db_xref="PID:g3874786"
/db_xref="GI:3874786"
/db_xref="SPTREMBL:Q93349"
/translation="MSLRAPRRSNADGLSENVLAILEPAAVMAETQMKKEGNG GSENTINDORNLOOASQSGISGFLDTLSLSTQSKVP"
20929..21011
/gene="C36F7.t1"
/note="TTG Gln Q-tRNA; predicted using tRNAscan-SE-1.11; preliminary prediction; similar to tRNA-Gln"
20929..21011
/gene="C36F7.t1"
24490..24574
/gene="C36F7.t2"
/note="TAT Ile I-tRNA; predicted using tRNAscan-SE-1.11; preliminary prediction; similar to tRNA-Ile"
24490..24574
/gene="C36F7.t2"
complement(25980..26064)
/gene="C36F7.t3"
/note="TAT Ile I-tRNA; predicted using tRNAscan-SE-1.11; preliminary prediction; similar to tRNA-Ile"
complement(25980..26064)
/gene="C36F7.t3"
gene

gene complement(27629..30314)
/gene="C36F7.3a"
complement(27629..30314)
/gene="C36F7.3b"
complement(join(27629..27714,28831..28875,29054..29189,29371..29522,30239..30314))
/gene="C36F7.3b"
/note="similar to IG (immunoglobulin) superfamily; cDNA EST YK427e11.3 comes from this gene"
/codon_start=1
/protein_id="CAB02819.1"
/db_xref="PID:e1344528"
/db_xref="PID:g3874791"
/db_xref="GI:3874791"
/translation="MSEYLCAVANGIPPEDESMYKLVTPPEPLVQAQSEYVQSGM ARMVCTEAMPREPMKDEGEPEYSENNVAMTIVGQYHSVHILEIRNVOSSHFCV YRCVAKNDNGIHHSQVTLNQSINHFTNSNLIPEELRAIQVDSASFSLITITIV YHGF"
complement(join(27629..27714,28185..28280,28831..28875,29054..29189,29371..29522,30239..30314))
/gene="C36F7.3b"
/note="similar to IG (immunoglobulin) superfamily; cDNA EST YK505g3.3 comes from this gene"
/codon_start=1
/protein_id="CAB02818.1"
/db_xref="PID:e1344527"
/db_xref="PID:g3874790"
/db_xref="GI:3874790"
/translation="MSEYLCAVANGIPPEDESMYKLVTPPEPLVQAQSEYVQSGM ARMVCTEAMPREPMKDEGEPEYSENNVAMTIVGQYHSVHILEIRNVOSSHFCV YRCVAKNDNGIHHSQVTLNQSINHFTNSNLIPEDESEDSQSPQLEFYQARHRS RNAPTRRELRAIQVDSASFSLITITIVYHGF"
complement(34255..36514)
/gene="C36F7.4"
complement(join(34255..34365,34413..34553,34617..34732,35806..35897,36303..36514))
/gene="C36F7.4"
/note="Similarity to Drosophila amagm protein (SW:AMAL_DROME); cDNA EST YK427e11.5 comes from this gene"
/codon_start=1
/protein_id="CAB02817.1"
/db_xref="PID:e1344526"
/db_xref="PID:g3874789"
/db_xref="GI:3874789"
/db_xref="SPTREMBL:Q93350"
/translation="MCVSSSEFFSSSTLPYFSRLHCLFLGKWTARSDALSAPFR GRGRFVILFHPIQERKMYLFALLGCVFKIPEARSPRLSVAFIATPRERFVFLKS CSRGAPPTIOQPSMSASVALIGQVDVCTCVNLDGSMVAFVADSPRLSDEKVF RRRNKYELKPRIGDLHNEWVLTIKNVOESDRGNTSCQINTEPITLSTGLDVKGIPFC RINRS"
complement(36940..38161)
/gene="C36F7.5"
complement(join(36940..37029,37077..37211,37260..37768,37839..37992,38081..38161))
/gene="C36F7.5"
/note="Predicted using GeneFinder"
/codon_start=1
/protein_id="CAB02815.1"
/db_xref="PID:e1344524"
/db_xref="PID:g3874787"
/db_xref="GI:3874787"
/db_xref="SPTREMBL:Q93351"
/translation="MLHMSFFYIFFTSCATIOVFLGKKKSGSKLSENSASTKS GKTGKSDKGESEPTPASKADPAPAPAPAPAAHPDIPYOTLAGINDVPEFKKGG AGGYGAPAPNARKGMAATHDYVOTLAGLNNVDFKKGGAARADGGGAPAADONA KAATHDPNVTLAGLNNVDFKKGGAARADGGGAPRAATGKRAANITKAAATHDYVOTLAG LNNVDFKKGGAARADGGGAPKPLDPPMKNQAAATLDIPYOTLAGVGDVFGADKKV GGAAGPGRKRPIDPPNKKGAATMDIPYOTLAGVGDVFGADKKR"
join(38494..38586,38728..38869,38969..39211,38325..1:109..380,283225..1:429..563,283225..1:834..923)
/gene="F4ZH11.1"
/codon_start=1
/protein_id="CAB02820.1"

NEIGHBORING COSMID INFORMATION:

The 3' cosmid is M01H9, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELR08C7; actual end is at 38112 of CELR08C7

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers

1..38112

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="R08C7"

1277..5453

/gene="R08C7.6"

join(1277..1396,1737..1890,2268..2785,3313..3850,

5065..5335,5381..5453)

/gene="R08C7.6"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48075.1"

/db_xref="PID:g1397321"

/db_xref="GI:1397321"

/translation="MRLIFVFIYVGVVGLGNSNPITNSGTPIDFGKAKCEAIIIG
DICTITFRVLETDNDANEFCRLVAPRLRLSRIRNEDGCIPOITCEVATMACDDGMI
OMEGYCFRHPYKNIKLTQAEAKVCSNLDGHTGETALHHRITGWMKNTRIGDQIH
VNGSTWNOYVSTGAVGDALALALHTGVAVSVTPSLIRIDPILRLGLCDKRE
RTLAIAVLGRVSEIYHQSIFLEDGLVLRSHVTSRTNAEICOKTLRPFVDVE
PIPDNDMSALSDQYTEPMALISNGAHTMDVSKRSSDCLSTHRSVAYDPPKM
PEVLPEVSVGCDNMNSVAIYHSEIDTEFRPMDSOSVPMCKEGRMNFVDPKG
YTHYRANGMHYGHVYSMILPEPKAMKCCDDGAGLGGFNDAEFAIDKLISKY
HWMIGRRRPPDCPGFTTSDPHNCGHKKHMLNVAETDPIELMFKRPDIAKMP
QGRQCGSIAYGMPGSNEPKIDGSCVKMPEICTAKGVLAATTASHL"

/complement(6651..8720)

/gene="R08C7.7"

complement(join(6651..6760,6809..6986,7091..7228,

7623..7764,7867..8168,8520..8720))

/gene="R08C7.7"

/note="similar to C. elegans olfactory receptor Odr-10
(GB:049449)"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48081.1"

/db_xref="PID:g2911717"

/db_xref="GI:2911717"

/translation="MHLSEFNNIPLKIVQYTSAPNSVSLICILIFLIAHSEPKLIGMT
KYLMIYISSEFALYSLMDVTETPMVHSYKAAVVRNENSDPREHSFILLIVYCGC
FGSLAIFGVHVEYRGAVVEKEDKMLGKRIYILFEMPIFYGTWMSVLYLVHFD
DSIDDMKRTIDGVDTKIEDTSYILFHPVDKGTSHDPAPASJACMFMFLSS
VFSVPEPGIKCVKRTIEALSRICNSVATKGLQDLFOALVQGFPIPLIMIPAIL
FRPMALVDIGRASSFVAMTIVYPAIDPLPMFIIKTRKRVAFSAIRNPGCSI
RFAQCTQIELVKSETPKNIGL"

8632..11285

/gene="R08C7.5"

join(8632..8864,9185..9226,9273..9384,9432..9581,

9635..9757,9980..10558,10623..10730,10777..10902,

10949..11164,11214..11285)

/gene="R08C7.5"

/note="similar to other C. elegans chemoreceptors"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48071.1"

/db_xref="PID:g1397317"

/db_xref="GI:1397317"

/translation="MRIQENETDINAENVCTTFRKTFONSLSKCIYASTATDGTITV
VAVGMAESEHIVHVEDMDGHRNGENKNNKNNFNQPHRISQIEYIKTAVFRSS
SRNHTKSDVAGAPTWDEFLKVDGMSQKLOECENQEKASNKVFALFKLNDVARKY
ATDPAEQLOKLIDRNYKPKPGFKDVKVEIEIAEHDELIFTAMYLSDASGSEITM

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

Query Match	13.1%;	Score 131;	DB 19;	Length 340801;
Best Local Similarity	82.1%;	Pred. No. 5.67e+49;		
Matches 238;	Conservative 0;	Mismatches 41;	Indels 11;	Gaps 8

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone

BASE COUNT 9566 a 5644 c 5839 g 9473 t
 ORIGIN
 Query Match 13.1% Score 131: DB 21: Length 30522;
 Best Local Similarity 81.2% Pred. No. 5.67e-49;
 Matches 229; Conservative 0; Mismatches 44; Indels 9; Gaps 5;

Db 12857 AACATAATCTGAGATGCGTTTGTGCAACACATTGACGGCAAAATATCTGTAGC 12916
 |||||
 Oy 3121 AACATAATCTGAGATGCGTTTGTGCAACACATTGACGGCAAAATATCTGTAGC 3180

Db 12917 GAAACACACATTAATGCTTAATGAGTAGCGCTGTGTGATTTAGCGGCTTAATTC 12976
 |||||
 Oy 3181 GAAACACACATTAATGCTTAATGAGTAGCGCTGTGTGATTTAGCGGCTTAATTC 3239

Db 12977 GAAACACATTTGCG--TAGTTTTTGTGCTAAC--TTT--TTTGCAATTTGATTTTCT 13030
 |||||
 Oy 3240 GAAACACATTAATGCTTAATGAGTAGCGCTGTGTGATTTGATTTTGTGTCTTCT 3239

Db 13031 TTTCTTATTTT--ATCGATTATAAATGATTTCCGTAATGACACAGCGCTACAGCAG 13088
 |||||
 Oy 3300 TTTGATATTTTGTGATCAATTAATTAATTTATTTCCGTAACAGACACAGCGCTACAGTAC 3359

Db 13089 TCACCTTAAGAGCTTACTGTAGCTTCCGTGTAAGAGATTTTG 13130
 |||||
 Oy 3360 TCTTTTAAGAGCTTACTGTAGCTTCCGTGTAAGAGATTTTG 3401

RESULT 11
 LOCUS AC006903 274626 bp DNA HTG 24-FEB-1999
 DEFINITION Caenorhabditis elegans clone Y74C10, WORKING DRAFT SEQUENCE, 15
 unorderd pieces.
 ACCESSION AC006903
 NID 94263445
 VERSION AC006903.1 GI:4263445
 HTG: HTGS-PHASE1.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
 Rhabdilitina; Rhabdilitidae; Rhabdilitinae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 274626)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Caenorhabditis elegans clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 274626)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (24-FEB-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2552: contig of 2552 bp in length
 * 2553 2567: gap of unknown length
 * 2568 5380: contig of 2813 bp in length
 * 5381 5395: gap of unknown length
 * 5396 8631: contig of 3236 bp in length
 * 8632 8646: gap of unknown length
 * 8647 12700: contig of 4054 bp in length
 * 12701 12715: gap of unknown length
 * 12716 18221: contig of 5506 bp in length
 * 18222 18236: gap of unknown length
 * 18237 23896: contig of 5660 bp in length
 * 23897 23912: gap of unknown length
 * 23912 31012: contig of 7101 bp in length
 * 31013 31027: gap of unknown length
 * 31028 38962: contig of 7935 bp in length

38963 38977: gap of unknown length
 * 38978 46049: contig of 7072 bp in length
 * 46050 46064: gap of unknown length
 * 46065 61946: contig of 15882 bp in length
 * 61947 61961: gap of unknown length
 * 61962 84927: contig of 22966 bp in length
 * 84928 84942: gap of unknown length
 * 84943 99867: contig of 15025 bp in length
 * 99868 99882: gap of unknown length
 * 99883 136628: contig of 36646 bp in length
 * 136629 136643: gap of unknown length
 * 136644 176604: contig of 39961 bp in length
 * 176605 176619: gap of unknown length
 * 176620 274626: contig of 98007 bp in length.

FEATURES
 source
 1. 274626
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="Y74C10"

BASE COUNT 87518 a 49452 c 49335 g 88111 t 210 others
 ORIGIN
 Query Match 13.1% Score 131: DB 19: Length 274626;
 Best Local Similarity 82.1% Pred. No. 5.67e-49;
 Matches 238; Conservative 0; Mismatches 41; Indels 11; Gaps 7;

Db 61526 CAAATAATCTGAGATGCGTTTGTGCAACACATTGACGGCAAAATATCTGTAGC 61585
 |||||
 Cp 3401 CAAATAATCTGAGATGCGTTTGTGCAACACATTGACGGCAAAATATCTGTAGC 3342

Db 61586 CGATTACGAGATCAATCT--TTAATCGAT--AAATATTTAAAGAAACACAAAAATGAC 61642
 |||||
 Cp 3341 CTGTTTACGAAATTAATTTATTAATGATCAAAAAATATCAAAAAACAAAAATGAC 3282

Db 61643 AAATAGCTTGAATATGCTGTGATTTGGAATATATCTGTTTGAATCCAGCCCG 61702
 |||||
 Cp 3281 AAATTTAA--AT--TCGTTGTCAAAATTCGAGATATATTCGTTTGAATCCAGCCCG 3226

Db 61703 TAAATGACCTACAGTAATAGTCAATTAAGAAATAGTCAATTTGCTTCCGCTACGAGATAT 61762
 |||||
 Cp 3225 TAAATGAC--ACTACAA--TAGTCATTTAAGAGGTACGTATTTTCCGCTACGAGATAT 3169

Db 61763 TTTGCGGCTCAAAATATGTTGTGCAATAGCGATTTTCCGATTCAGTATTC 61812
 |||||
 Cp 3168 TT-GCGGCTCAAAATATGTTGTGCAATAGCGATTTTCCGATTCAGTATTC 3120

RESULT 12
 LOCUS AC006751 340801 bp DNA HTG 23-FEB-1999
 DEFINITION Caenorhabditis elegans clone Y93G10, WORKING DRAFT SEQUENCE, 29
 unorderd pieces.
 ACCESSION AC006751
 NID 94263187
 VERSION AC006751.1 GI:4263187
 HTG: HTGS-PHASE1.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
 Rhabdilitina; Rhabdilitidae; Rhabdilitinae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 340801)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Caenorhabditis elegans clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 340801)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

NOTE: This is a 'working draft' sequence. It currently
 * consists of 29 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

Db 192245 GATTTCGAAATGATTTATTTTCGAAATGACACGAAATTTTCTTACTTTT 192304
 QY 3233 GATTTCGAAATGATTTATGCTGAAATGTCACAGAA-TTTTATTTGCTTTTGG 3291
 Db 192305 TGTGTTTCTTTTAAATTTTATTCATTAATGATTTCCGTAATGACACGACGCC 192364
 QY 3292 TGTGTTTCTTTTGAATTTTATTCATTAATTAATTTTCCGTAATGACACGACGCC 3351
 Db 192365 TACAGTAGTATTCAGAAATTTACTGATTTTCTACTAGACGATTTTGG 192414
 QY 3352 TACAGTAGTATTTTAAAGTATGATTTTCTGCTTCAAGATTTTGG 3401

RESULT 10
 LOCUS CELM01H9 30522 bp DNA INV 30-SEP-1996
 DEFINITION Caenorhabditis elegans cosmid M01H9.
 ACCESSION U70853
 NID 91572795
 VERSION 070853.1 GI:1572795
 KEYWORDS
 SOURCE
 ORGANISM
 Caenorhabditis elegans strain-Bristol N2.
 Eukaryote; mitochondrial eukaryotes; Metazoa; Nematoda;
 Secernitea; Rhabdita; Rhabditida; Rhabditina; Rhabditoidea;
 Rhabditidae; Caenorhabditis.
 1 (bases 1 to 30522)
 Wilson, R., Alnscough, R., Anderson, K., Baynes, C., Berks, M.,
 Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
 O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, N.,
 Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
 Wilkinson-Sproat, J. and Wohlman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT
 Nature 368 (6466), 32-38 (1994)
 94150718
 2 (bases 1 to 30522)
 Pauley, A and Gattung, S.
 The sequence of C. elegans cosmid M01H9
 Unpublished (1996)
 3 (bases 1 to 30522)
 Waterston, R.
 Direct Submission
 Submitted (15-SEP-1996)
 Submitted by:

Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RO, England
 e-mail: rtwenematode.wustl.edu and jesusanger.ac.uk

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is R08C7, 200 bp overlap; 3' cosmid is ZK180, 200 bp
 overlap. Actual start of this cosmid is at base position 92 of
 CELM01H9; actual end is at 11962 of CELZK180

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program GeneFinder (P. Green and L. Hillier, ms in preparation).
 Location/Qualifiers

FEATURES
 source
 1..30522
 /organism="Caenorhabditis elegans"

gene
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="M01H9"
 complement(1636..3073)
 /gene="M01H9.2"
 complement(join(1636..1850,1909..2237,2298..2560,
 2609..2685,2738..2872,2923..3073))
 /gene="M01H9.2"
 /note="weak similarity to mouse fibinlin-1 precursor
 (GI:575510) and the Emericella nidulans anucleate primary
 sterigmata coiled-coil protein (GB:X8289)." /evidence=not_experimental
 /codon_start=1
 /protein_id="AAB09143.1"
 /db_xref="PID:91572797"
 /db_xref="GI:1572797"
 /translation="MQLDLFLIFCSFTNASSCQDPSYCNPNDCVPRGVAMVYCRK
 TGNCADFCEDSKFTCSARKKDCDDMSDPCPRGRCYAKLKPDSTRKTIPTVQ
 FKNRPATSTTTTTRSPSIRMKQPMPLNGTFINPLPKYEKLDITTYIPEHTVI
 DYPIDAMEELIMEPORIMPEPEPMRLEPISYSPYVPOVPOYSPPSPILGI
 SSSLDSSRRAPQWESSYGOYPOYDEKINTIAPNMQVPEFLIPGQNDLPKMSMS
 LNLICXKDKDEVICKHVTRDCLSPRTYLLDPCYCNKCSGYOCIDSKIDCKAEVK
 AGACGLSVASEICPRTCEYCNPPSDMARTMSDCKDELETCDQLAESGACQHDPR"
 complement(9944..11767)
 /gene="M01H9.3"
 complement(join(9944..10045,10099..10309,10362..10510,
 10745..10868,11739..11767))
 /gene="M01H9.3"
 /note="coded for by C. elegans cDNA yk151a7.5; coded for
 by C. elegans cDNA yk18419.5; coded for by C. elegans cDNA
 yk91a11.5; coded for by C. elegans cDNA yk99b12.5; coded
 for by C. elegans cDNA yk91a11.3; coded for by C. elegans
 cDNA yk151a7.3"
 /codon_start=1
 /protein_id="AAB09144.1"
 /db_xref="PID:91572798"
 /db_xref="GI:1572798"
 /translation="MPLDPMDSLMASKAPQVSDLVVOPNVTFKAVTGSPSPSASIP
 IAGADSIIPDYDTCMLKPPTRSAVROKERTISDSDSEIGALGGRKFTNEGASIP
 APTSPIDAOPTARRMSISEMLEGSSPKSPKSGMSENTNTTTSNGSGDGIGERKMSI
 TDVSGMKQKQIDPRKDPKHKOSKIIGDGIASAGFKRSNTMKD"
 complement(20695..23630)
 /gene="M01H9.4"
 complement(join(20695..20957,21298..21624,22484..22621,
 22670..22751,23583..23630))
 /gene="M01H9.4"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB09145.1"
 /db_xref="PID:91572799"
 /db_xref="GI:1572799"
 /translation="MSEFVSLKRPSESDVTEEPDEKYSIHAGNHQOLMNPAYMAG
 PNARGLQNVGPPLMPOYCOMIQOHQHQQEYEPMPTRKNDKDEANSASTPSSSR
 QTLSSGAGPRIRIVPOSITNPQGMPLSQOQIIEEPTSPITMRQOSDHTYH
 MFSFORADEEITYVAIFEDFISLRGRKIKTOSIFRGHKTPPIRFPSNMISTL
 FSVIPILVWELVFGVLAPDEVMDVFGENVPIDNRIVSPQSVADVLSGDCG
 RGGELAVLS"
 28085..29825
 /gene="M01H9.1"
 join(28085..28168,29021..29086,29201..29494,29745..29825)
 /gene="M01H9.1"
 /note="contains weak similarity to thioresoxin
 redox-active center"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB09142.1"
 /db_xref="PID:91572796"
 /db_xref="GI:1572796"
 /translation="MAKNFESGTLRLKDGTMVDAGELHKGRIVLVYSASWGPCRO
 FTPIKLELYOOIATNOPILFYLSDRYMEQDVEYSGSCGVSVPALDPIREKT
 EKDYVKALPSCRVVDFEGNCIDANAROSVEVNIENNRNVPESREYREKYMAELF
 NKWTEKGNLEIRIA"

[illegible]

D6 2151 TTTCGCGCTCAAAATGTGTGGCAGTACGTATTCTCAGAATTTTTCGCTC 2200
|| ||||||| |||| | | || ||||||| ||||
Cp 3168 TT-GCCGCTCAAAATATGTTGAGTAATGCCCATTTCTCAGAAATTTACTGTTTC 3120


```

/db_xref="PID:93880758"
/translation="MSATEKDKTPPLGVSQPSISLAPDSDKDIAOTLTLLIETLKSSG
EPKKEEQRMEVLNRLNRLVKNVKNNTAKITNGBNVGGKLTFFGSGRLGVHSSG
ADIDTLAVPRHIDRSDFTSPKEMLNNDNVLEGEVAEPVAKLTSVGLDIL
FARLALKEVDPTOLSDNDLRLNDLDOESVSLNGCRPAEDLTLVROKFECTTLAI
KLAKNGIYSNSMGFGGILMALIVACOLVSNVSRHLVHMFIFISWTWPHV
VLENNNDNDIPICELVMDPRKRNDRFHPVHIIIPAPENSTHNTVSTAYIK
NTEFALEICDISEGSKTALFEEPNRSPKAKDYNENKRNFEGLFAPLAAATL
RIPLVSLERNODIILAHNDPNKRSKPNKADYNENKRNFEGLFAPLAAATL
DLTNEIORKTNVLEASNVKIGPNCQVIDHYKRNLIQVTSNADIRRRNRK
VPIATINVSSTSPRSVTRTSTSVPTPTGLAARPLSLASVSATNPDSTINGT
PLSRKRSDEESTTTSQISDESVPKRTDREBNRSMVSVNVEVDEQRTVY
QEIYDLADNDGINTNGLESECKEYPOSV"
join(complement(271180,1:2698, .208),
complement(271180,1:1785, .2041),
complement(271180,1:1541, .1741),
complement(271180,1:1152, .1497),
complement(271180,1:115, .1106),complement(27038, .27236),
complement(26919, .26986),complement(26026, .26313),
complement(25704, .25934),complement(25543, .25653),
complement(24197, .24835),complement(19242, .19310))
/gene="F2E12.1"
/notes="similar to BPTI/KUNITZ inhibitor domain. cDNA EST
EMBL:D68293 comes from this gene; cDNA EST YK448H4.5 comes
from this gene; cDNA EST YK42496.5 comes from this gene;
cDNA EST YK448H4.3 comes from this gene"
/codon_start=1
/protein_id="CA16311.1"
/db_xref="PID:9380302"
/db_xref="PID:93880760"
/db_xref="GI:3880760"
/db_xref="SPTREMBL:045916"
/translation="MLTVLLIFGCTAAVNSQQLSNPKCNHMPRGTCELDFHVKY
DVRDRCRFFGCGEENNRDLSLECSOCHOVPTNRDRCOPHPDPCYADIR
WFFDENKQOCVSMNSCGGNSNYYSNHLMLGAEHGPIDEXKMGKRNRM
SAESLHVENPTESTHOTSSEPTPOVPLENSYDHSRRRNGSDNDRKMTINIS
RDGSGVFNPAATVPMFSRAOSFKTQADSLTHRTDSEPRPQVLDVQVNOIOE
QASAOQNGMRRKFRKOKKLPRLMHLGTRPASOSNSYRQOLSENEHRVPTPO
VREODAHVOQHNOISQSLONEAETORIODMRNVEPEOAROPVPMPPPPPO
ONSPVNEYHPELVNOFSLDEHEKARLELAOPDHLITLPHLETVRNDSN
VYRQKRTKAYPKGARIPHYVPPVAPLPPMPQOVIOTPLTOAVTYEPIPL
ITVPHETISRLKHEMOKKYKEDLEKROEQRKROEHEKRLAAGQREBRM
VEREKAENDRRLKTEENKQAEYQARMLTITPTTYETTYEYHYVPIPLPV
EYVYDAHKRPIPTSKSLAEPEHFRPIIENVTASLSIEDIITANDRSTEDEDY
PMPFDEAPPLTPRPVYNTTRPPPPQIPQPSRINGMPSIVQPSGGKSKL
FAPSESEYTOEDEDPYNLLFLLSLAAGASNSVNGVLYOYDEDMGYVDIR
VADMYVNDPVNGESYVDDKSAVTEITQIDKLAMLRKLLALHPLEPRPEYIFC
ESTGKMSAQFAVITDSTKASLOLNFAVALGDSVISAADYNTMGPVANSFEL
DHOJLNTVNEAARQALVDOOKWEKATNCWGMENLJSEIDVSNYINLLKGDIDM
SSSAMSNSVNTSTRLYNRFPAPOLDSLNYMDTVYKRGKGIIPDYKRGOGADV
FSYOGDFMPTIWSYDOLKDGYNIVYNGNEDILCNMGTAANRLTMGATFN
STTRHSFKTOSPLAGYKYTKNLOFWMILRAGHMAVATPESAFEMKAYVKKQNA"
join(complement(6241, .6293),complement(6089, .6195),
complement(5656, .5745),complement(5110, .5281),
complement(4965, .5062),complement(4090, .4182),
complement(2542, .2748),complement(2003, .2488),
complement(1841, .1944),complement(1734, .1787),
complement(723, .788),complement(325, .670),
complement(105, .141),complement(273972,1:25167, .25503),
complement(273972,1:25055, .25113),
complement(273972,1:24908, .25008),
complement(273972,1:24744, .24861),
complement(273972,1:24468, .24612))
/gene="F15H10.4"
join(complement(6241, .6293),complement(6089, .6195),
complement(5656, .5745),complement(5110, .5281),
complement(4965, .5062),complement(4090, .4182),
complement(2542, .2748),complement(2003, .2488),
complement(1841, .1944),complement(1734, .1787),
complement(723, .788),complement(325, .670),
complement(105, .141),complement(273972,1:25167, .25503),
complement(273972,1:24468, .24612))

```

... Note: remainder of annotations omitted.

Query Match 13.4%; Score 134; DB 21; Length 27236;
 Best Local Similarity 83.2%; Pred. No. 1,51e-50;
 Matches 243; Conservative 0; Mismatches 37; Indels 12; Gaps 9;

```

Db 8976 AAAATATCTGTAAGCAAACTAGCACTATTTAAATGACTACTGATGCGCTGCTC 9035
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3400 AAAATATCTGTAAGCAAACTAGCACTATTTAAATGACTACTGATGCGCTGCTC 3341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9036 GATTTCGGAATATCTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 9095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3340 TGTTCAGGAATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9096 ACGAAATTAATATTTGCTGCGCAATTCGCAATTAATTAATTTAGTTTATGAGCCG 9155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3283 AC--AAATTAATTTGCTGCGCAATTCGCAATTAATTTAGTTTATGAGCCG 3226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9156 TAAATGACACTAGCTTACATTTAGCATTTAAGAAATGATGCTGCTGCTGACACA 9215
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3225 TAAATGACACTA-C--A-AT-AGTCATTTAAGGCTTACTGATTTTTCCTACGAGA 3172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9216 TACTTTCGCGCTCAATATGCTATGATTTAGCATTTGCAATTTAGTGT 9267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3171 TA-TTTCGCGCTCAATATGCTATGATTTAGCATTTGCAATTTAGTGT 3121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 5
 LOCUS CER07H11 36879 bp DNA HTG 14-OCT-1998
 DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
 clone K07H11, WORKING DRAFT SEQUENCE.
 ACCESSION Z81566
 NID 93377978
 VERSION Z81566.1 GI:3377978
 HTG: HTGS_PHASE1.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
 1 (bases 1 to 36879)
 Steward, C.
 Direct Submission
 Submitted (14-OCT-1998) Nematode Sequencing Project, Sanger Centre,
 Hinxton, Cambridge CB10 1HQ, UK and Department of Genetics,
 Washington University, St. Louis, MO 63110, USA. E-mail:
 joesanger.ac.uk or trhematode.wustl.edu

COMMENT

On Aug 3, 1998 this sequence version replaced gi:1695068.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n's separate
 segments.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source location/Qualifiers
 1..36879
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="V"
 /clone="K07H11"
 BASE COUNT 11032 a 6897 c 6237 g 11112 t 1601 others
 ORIGIN

Query Match 13.4%; Score 134; DB 19; Length 36879;
 Best Local Similarity 83.2%; Pred. No. 1,51e-50;
 Matches 243; Conservative 0; Mismatches 37; Indels 12; Gaps 9;
 Db 14420 AAAATATCTGTAAGCAAACTAGCACTATTTAAATGACTACTGATGCGCTGCTC 14479

Cp	3460	GAATAACCGCTTTGGCCACATGTTAGATGATTTTTTTTTCAAAGATGTTAAATCTTTTC	3401
Db	9954	AAATATCTTGAAGCGAAACACTACTGTAACCTTTTAAAAAGTACTGTAGCGCTGTGTC	10013
Cp	3400	AAAATATCTTGAAGCGAAACACTACTGTAACCTTTTAAAAAGTACTGTAGCGCTGTGTC	3341
Db	10014	TGTTTAGCGAATAATTAATTAATGATCAAAATATCAAAAGAAACCAAAATATGCA	10073
Cp	3340	TGTTTAGCGAATAATTAATTAATGATCAAAATATCAAAAGAAACCAAAATATGCA	3281
Db	10074	AATTAATAATTCGTGTCACAATTCGAGCATATATTCGTTTGAAAAATCGACCCGTAAAT	10133
Cp	3280	AATTAATAATTCGTGTGTCACAATTCGAGCATATATTCGTTTGAAAAATCGACCCGTAAAT	3221
Db	10134	CGACACTCAATAGTCATTTTAAAGGGTACTGTAATTTTTCCTACGAGATATTTTGGCGG	10193
Cp	3220	CGACACTCAATAGTCATTTTAAAGGGTACTGTAATTTTTCCTACGAGATATTTTGGCGG	3162
Db	10194	TCAATATGTTGAGTAATGCGCATTTCTAGAATTTAGTGTTCGCATCAACATACAAAGTT	10253
Cp	3161	TCAATATGTTGAGTAATGCGCATTTCTAGAATTTAGTGTTCGCATCAACATACAAAGTT	3102
Db	10254	GGCTGGAAATGCGGTGACTGGAGAGATGATATATTTGTGCGATCCGATGAATGAAGTG	10313
Cp	3101	GGCTGGAAATGCGGTGACTGGAGAGATGATATATTTGTGCGATCCGATGAATGAAGTG	3042
Db	10314	CACGGATCGAAGACGAGATCTTGCTCTTGTGATAGATATCTCGATTAAGAATGATGATG	10373
Cp	3041	CACGGATCGAAGACGAGATCTTGCTCTTGTGATAGATATCTCGATTAAGAATGATGATG	2982
Db	10374	ACACTGAAGAGACGCTGTTCAGGGTAACTCGGGTGGTGAAGTGTAGCGGGCGGGCTCA	10433
Cp	2981	ACACTGAAGAGACGCTGTTCAGGGTAACTCGGGTGGTGAAGTGTAGCGGGCGGGCTCA	2922
Db	10434	ATGCGCGGCTCGAGATGCGCTTGCCGGTGAATGGACACTCGAATTCGACAGCATTCG	10493
Cp	2921	ATGCGCGGCTCGAGATGCGCTTGCCGGTGAATGGACACTCGAATTCGACAGCATTCG	2862
Db	10494	AGTCAACACTGGAATAAAGGATATTTGAAGAGATGTGTGTGTGTGTGTGACAACTA	10553
Cp	2861	AGTCAACACTGGAATAAAGGATATTTGAAGAGATGTGTGTGTGTGTGTGACAACTA	2803
Db	10554	TTTTTGGCAATCTTCAAAATATGCGCTTTTGGCCACGAAATTAAGTGTAAAAATAAT	10613
Cp	2802	TTTTTGGCAATCTTCAAAATATGCGCTTTTGGCCACGAAATTAAGTGTAAAAATAAT	2743
Db	10614	AATCACTGAGCTTCAGTAATTAATAATATATTCACTTCAA	10655
Cp	2742	AATCACTGAGCTTCAGTAATTAATAATATATTCACTTCAA	2701

RESULT	3	CEY67H2	314495 bp	DNA	HTG	04-MAR-1999
LOCUS		Caenorhabditis elegans	DNA ***	SEQUENCING IN PROGRESS ***	from	
DEFINITION		clone Y67H2, WORKING DRAFT SEQUENCE.				
ACCESSION		AL022475				
NID		94469034				
VERSION		AL022475.3	GI:4469034			
KEYWORDS		HTG, HTGS, PHASE1,				
SOURCE		Caenorhabditis elegans.				
ORGANISM		Caenorhabditis elegans				
		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;				
		Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.				
REFERENCE		1 (bases 1 to 314495)				
AUTHORS		McMurray, A.				
TITLE		Direct Submission				
JOURNAL		Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,				
		Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,				
		Washington University, St. Louis, MO 63110, USA. E-mail:				
		jesssanger.ac.uk or twenematode.wustl.edu				
COMMENT		On Mar 22, 1999 this sequence version replaced gi:4468145.				
		IMPORTANT: This sequence is unfinished and does not necessarily				
		represent the correct sequence. Work on the sequence is in progress				

```

FEATURES
    source
        1..314495
        /organism="Caenorhabditis elegans"
        /db_xref="taxon:6239"
        /chromosome="IV"
        /clone="Y67H2"
BASE COUNT
965442 a 54551 c 52760 g 97027 t 13615 others
ORIGIN

```

Query Match	93.0%	Score 930	DB 19	Length 314495
Best Local Similarity	98.9%	Pred. No. 0.00e+00		
Matches 991	Conservative	0	Mismatches 1	Indels 10
			Gaps	9
Db 239319	AAATGATAGCGCTTTTATAC-AATTTTCTGCACA-CATGATAGGGGAAAATATATATAAA	239376		
Cp 3700	AAATGATAGCGCTTTTATACATTTTCTGCACACATGATAGGGGAAAATATATATAAA	3641		
Db 239377	AAT-GC-ATTTTTCGTAT-CATATGAACA-CCACATGATTTTTCGTGACCATTTTC	239432		
Cp 3640	AATGCTATTTTTCGTATTCATATGAAACACACATGATTTTTCGTGACCAATTTTC	3581		
Db 239433	AAAAGTAGACATCTATGAACGAAAA--CCAAATTTCTCAAGATATTCACATTAATGA	239490		
Cp 3580	AAAAGTAGACATCTATGAACGAAACCCAAATTTCTCAAGAAATTCACATTAATGA	3521		
Db 239491	TAAATATATGTGTTAATATAGCATATCTTGATGAAAAAAATCAAAAAATTCGGAAACTT	239550		
Cp 3520	TAAATATATGTGTTAATATAGCATATCTTGATGAAAAAAATCAAAAAATTCGGAAACTT	3461		
Db 239551	GAAAAAACGTTTTGGACATGTTAGATGATTTTTCCAAATGTTTAAATTCCTTTC	239610		
Cp 3460	GAAAAAACGTTTTGGACATGTTAGATGATTTTTCCAAATGTTTAAATTCCTTTC	3401		
Db 239611	AAAATATCTTAAACGAAAACTACTGTAACCTTTTAAAGACTGTAGCGCTGTGTC	239670		
Cp 3400	AAAATATCTTAAACGAAAACTACTGTAACCTTTTAAAGACTGTAGCGCTGTGTC	3341		
Db 239671	TGTTTACGGAATATATTTATTAATGATCAAAAATATCAAAAGAAAAACAAAAATGACA	239730		
Cp 3340	TGTTTACGGAATATATTTATTAATGATCAAAAATATCAAAAGAAAAACAAAAATGACA	3281		
Db 239731	AATTAATTTGTTGTGCACATTCGACATATATTTGTTTCGAAAAATCGAGCCCGTAAT	239790		
Cp 3280	AATTAATTTGTTGTGCACATTCGACATATATTTGTTTCGAAAAATCGAGCCCGTAAT	3221		
Db 239791	CGACACTCAATAGCATTTTAAAGGTTACTGTAATTTTCGCTACGAGATATTTGCGCG	239850		
Cp 3220	CGACACTCAATAGCATTTTAAAGGTTACTGTAATTTTCGCTACGAGATATTTT-GCGCG	3162		
Db 239851	TCAATATGTTGAGTAAATGCCCATTTTCACAAATTTAGTGTTCGATCAACATACAAAGTT	239910		
Cp 3161	TCAATATGTTGAGTAAATGCCCATTTTCACAAATTTAGTGTTCGATCAACATACAAAGTT	3102		
Db 239911	GCGTGGAAATCGCTGTGACTGGAGAGATGAATATGTTGTGATCCGATGAATGAAGTG	239970		
Cp 3101	GCGTGGAAATCGCTGTGACTGGAGAGATGAATATGTTGTGATCCGATGAATGAAGTG	3042		
Db 239971	CACGGCATCGAAACGAGATCTTCTCTTGATAGATATCTCTGATAAGAGTGAATGATG	240030		
Cp 3041	CACGGCATCGAAACGAGATCTTCTCTTGATAGATATCTCTGATAAGAGTGAATGATG	2982		
Db 240031	ACACGAGACACGCTGTACAGGGAACCTGGGTCGGGTGAAGTGAACGGCGGGGCTCA	240090		
Cp 2981	ACACGAGACACGCTGTACAGGGAACCTGGGTCGGGTGAAGTGAACGGCGGGGCTCA	2922		

REFERENCE
AUTHORS
2 (bases 1 to 39908)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Spratt, J., and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
Nature 368 (6466), 32-38 (1994)

JOURNAL
MEDLINE
94150718
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:
http://webcpc.sanger.ac.uk/cgi-bin/displaydb-wormace/class=sequence&object=C48D1
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
This sequence is the entire insert of clone C48D1. The true right end of clone F38D2 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence Z81093. The end of this sequence (33020..39908) overlaps with the start of sequence Z82274.

FEATURES
SOURCE
1..39908
location/Qualifiers
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(994..4439)
/gene="C48D1.1"
complement(join(994..1020,1345..1951,2283..2580,2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/gene="C48D1.1"
/note="similar to C. elegans zyx-11 like protein; cDNA EST EMBL:Z14909 comes from this gene; cDNA EST EMBL:J01768 comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="pid:el344741"
/db_xref="pid:g3875048"
/db_xref="gi:3875048"
/db_xref="sp:trEMBL:O02227"
/translation="MTKPLHDIACOKTAEVQNGSYDNREFELDSKSNIVSYLMNRDKPNLTKIKLNTKYLTHPOLIDENTARQLOKOFSLKELKIGNFSKSCWCFRRKETHVDIYRIEAVLNERTKRLQYLTHERKQKMLKQVLAIGMLPSITLHMVDVHINEDFOYLPNGNSRLSDISGTNIINIAISIKRQVLALEDEFATITTELDKLFN LKLRVLDVQSOTRNLHLSRNIVYBOFTCEMELPELHMCCTTSLSGDIIELECLTQ LSTPNIISVTEAEISQINDVRLVNTAFHNSVQALVYCLTARREFSGCKMKAIRISI SEDNQLNINPSCIKITVKSMTQFDSISVONGIKCLIFGMLYSQDSNGVTLTIVE ALMSFVLYKRMISSEAYKVNISYMNVAEKLMLKGLDFDKISKAMSFLNNFVS KALTRASSVLCOCIRHMSDELKSLSEKENVNRIYDILEMSVGOKNEMTKSEFEI TIKTSSKACATNMIOAGIISISLSCEMEVESNDITTKLELTITNLSALNDSLOE LITSNVDKIGQANAYSTER"
complement(6258..11120)
/gene="C48D1.2"
complement(join(6258..6349,6876..7065,7420..7547,7602..7815,8728..9059,10248..10501,10780..10925,10989..11120))

gene
CDS
complement(join(16940..17126,18419..18597,18985..19442,19533..19619,19925..19961,20333..20481,20789..20987,21037..21152,21373..21616))
/gene="C48D1.3"
/note="Similarity to Salmonella sodium/proline symporter (SW:PUTP_SALT); cDNA EST EMBL:C1840 comes from this gene; cDNA EST EMBL:C1562 comes from this gene"
/db_xref="sp:trEMBL:O02228"
/db_xref="pid:el344742"
/db_xref="pid:g3875049"
/db_xref="gi:3875049"
/translation="MADLIGYAIYFVLLVYGNAGRSKSKLESEAGATEE VLAGRNIGTLVIGTMTATWGAIVNGTAEALYNGLLCCQAPVGAISLVGGL FAKKMRREGVITMDLPFOFMNLELIFGRFDNFRKLGRLQIIEIDFPQHKY QRIIGLMLVPALIGDEFTWAAILSLMGLATLSVIGIDNNAVTLTASACIAYFTGG YAVATDVVDFECIFVGLLIGLYVONPNPKFSLIMDMLTIPVGNKIMVOYFQF VLSKTAGAOITLSPVAGCTIIMAPALICAIARNDMRTYSPNNKTKYESIP DKRMNVPYFQYITPRVAFIGLGAIVMSADSVSLSASMRHNHNTKLTIRP HNSEKEVITVNRALITGVIMATLALQSIYGLMGLADLVYVILFPLLCAVYMP KSTNYSLAGVAVGLVRLIGGEPLVSLPAFFHVPMTGDGVYVFTFLLCSMATI YVSIOSKEIKFSKRLSPEDMVGCVVMIPIDHVLPEDVFAVSEFLNNKVCDCGM DEPOLQTEHRIQYIRTNRLMKIHVIHFRKVIILPQAIKRDVLVPRVHLISSQIFF CTFESFR"
BASE COUNT 13182 a 6945 c 7146 g 12635 t
ORIGIN
Query Match 93.0%; Score 930; DB 21; Length 39908;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 991; Conservative 0; Mismatches 1; Indels 10; Gaps 9;
Db 9662 AAATGATCGCGCTTTTTCAC-ATTTTCGCACA-CATGATATGGGGAATAATATATAAA 9719
|||
Cp 3700 AAATGATCGCGCTTTTTCATATTTTTCGCACAACATGATGAGGGAATAATATATAAA 3641
|||
Db 9720 AAT-GC-ATTTTTCGCAT-CATATGACAA-CCACATGATTTTTTTCGTGACCAATTC 9775
|||
Cp 3640 AATTTGATATTTTTCGATTCATATGAAACCAACATGATTTTTCGTGACCAATTC 3581
|||
Db 9776 AAAAGTAGAGCATACAGAGAGAAAAA-CCAAATTTTACAGAGATTCACATTAATGA 9833
|||
Cp 3580 AAAAGTAGAGCATACAGAGAGAAAAAATTTTTCACAGAGATTCACATTAATGA 3521
|||
Db 9834 TAATTAATGTGTATAAAGCATATCTGAATGAATAAATAAATAATCTCGAAACTT 9893
|||
Cp 3520 TAATTAATGTGTATAAAGCATATCTGAATGAATAAATAAATAATCTCGAAACTT 3461
|||
Db 9894 GAAAAAGCGTTTGGCACATGTTGATGATTTTTTTTCAAAATGTTTAAATTCCTTTC 9953
|||

Query Match	Best Local Similarity	100.0%: Score	1000: DB	Length	7653:
Matches	1000: Conservative	0: Mismatches	0: Indels	0: Gaps	
Db 2701	TTGAAGTGAATATATATTATTATTACTGAAAGCTCGAGTATTATTATTATTAACTA	2760			
Db 2701	TTGAAGTGAATATATATTATTATTACTGAAAGCTCGAGTATTATTATTAACTA	2760			
Db 2761	ATTTCGTGGCGCAAAAGGCCATTGTTGATGATTTGCCGAAATACTTGTACACACACAC	2820			
Db 2761	ATTTCGTGGCGCAAAAGGCCATTGTTGATGATTTGCCGAAATACTTGTACACACACAC	2820			
Db 2821	ACACACATCTCTCAAAATATCCCTTTTCCAGTGTGACCTGGAATGCTGTGCAATTGCA	2880			
Db 2821	ACACACATCTCTCAAAATATCCCTTTTCCAGTGTGACCTGGAATGCTGTGCAATTGCA	2880			
Db 2821	ACACACATCTCTCAAAATATCCCTTTTCCAGTGTGACCTGGAATGCTGTGCAATTGCA	2880			

Db	2881	GTGTCATGTGCACGGCCAGCGCATGTGGAGACGGCGCATTTAGACCCGGCGGCTAC	2940
Qy	2881	GTGTCAATGTGCACCGGCAGGCCATGTGGAGACGGCGCATTTAGACCCGGCGGCTAC	2940
Db	2941	TTCAACGCAGCCGAGCTTACCGGTGACAGCGCTCTTTCAGTGTATCATCTTCCATTATCA	3000
Qy	2941	TTCAACGCAGCCGAGCTTACCGGTGACAGCGCTCTTTCAGTGTATCATCTTCCATTATCA	3000
Db	3001	GGATATCTACTCAAGACGAATCTGTTCTGATCGCGTGACATTCTATCGGATCG	3060
Qy	3001	GGATATCTACTCAAGACGAATCTGTTCTGATCGCGTGACATTCTATCGGATCG	3060
Db	3061	ACACAAATATTCATCTCTCCAGCTCAACGCATTTTCCAGCCAACTTATATGTTGATGCG	3120
Qy	3061	ACACAAATATTCATCTCTCCAGCTCAACGCATTTTCCAGCCAACTTATATGTTGATGCG	3120
Db	3121	AACACTAAATTTCTGAGATGCGCATTTACTCAACATATTTGACGGCCAAATATCTCGTAGC	3180
Qy	3121	AACACTAAATTTCTGAGATGCGCATTTACTCAACATATTTGACGGCCAAATATCTCGTAGC	3180
Db	3181	GAAGAAATACAGTAAACCTTTAAATGACTTTGTAGTGTGATTAAGGGCTCGATTTTTCG	3240
Qy	3181	GAAGAAATACAGTAAACCTTTAAATGACTTTGTAGTGTGATTAAGGGCTCGATTTTTCG	3240
Db	3241	AAACGAATATATGCTGCAATTTGACAAAGAAATTTTAAATTTGTCTATTTTGTGTTTCTT	3300
Qy	3241	AAACGAATATATGCTGCAATTTGACAAAGAAATTTTAAATTTGTCTATTTTGTGTTTCTT	3300
Db	3301	TTGATATTTTGTGATCAATTTATATAATTTTCCGTAAACAGACACACGCGCTACAGTACT	3360
Qy	3301	TTGATATTTTGTGATCAATTTATATAATTTTCCGTAAACAGACACACGCGCTACAGTACT	3360
Db	3361	CTTTTAAGAGGTACAGTATGTTTGGCTTCAGATATTTTGGAAAGAAATTTTAAACATTT	3420
Qy	3361	CTTTTAAGAGGTACAGTATGTTTGGCTTCAGATATTTTGGAAAGAAATTTTAAACATTT	3420
Db	3421	TGAAAAAAATTCATCTACATGTGCCAAACGCTTTTTCAGAGTTTGCACATTTTGTGA	3480
Qy	3421	TGAAAAAAATTCATCTACATGTGCCAAACGCTTTTTCAGAGTTTGCACATTTTGTGA	3480
Db	3481	TTTTTTTCATCAAGATATGCTTTATTAACACATATTAATATCATTAATGSAATTTCTTG	3540
Qy	3481	TTTTTTTCATCAAGATATGCTTTATTAACACATATTAATATCATTAATGSAATTTCTTG	3540
Db	3541	TAGAAATTTTGGGCTTTGCTCTAGTATGCGCTACTTTTGAATTTGCTACAGCAAAAAA	3600
Qy	3541	TAGAAATTTTGGGCTTTGCTCTAGTATGCGCTACTTTTGAATTTGCTACAGCAAAAAA	3600
Db	3601	TCATGTGTTTGTTCATATGAATGACAAAAATAGCAAAATTTTATATATTTTCCCTAT	3660
Qy	3601	TCATGTGTTTGTTCATATGAATGACAAAAATAGCAAAATTTTATATATTTTCCCTAT	3660
Db	3661	TCATGTGTGCGAAAAATAGTAAAAAAGCCGATCAATTT 3700	
Qy	3661	TCATGTGTGCGAAAAATAGTAAAAAAGCCGATCAATTT 3700	

RESULT	2						
LOCUS	CEC48D1	39908 bp	DNA	INV	23-NOV-1998		
DEFINITION	Caenorhabditis elegans cosmid C48D1, complete sequence.						
ACCESSION	281049						
NID	91627677						
VERSION	281049.1	GI:1627677					
KEYWORDS	HTG.						
SOURCE	Caenorhabditis elegans.						
ORGANISM	Caenorhabditis elegans.						
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Pelodertinae; Caenorhabditis 1 (bases 1 to 39908)						
AUTHORS	Burton, J.						
TITLE	Direct Submission						
JOURNAL	Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jesse@sanger.ac.uk or iw@nematode.wustl.edu						

 WISEVIEW (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrcrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 04:15:39 1999; MasPar time 1830.49 Seconds
 1514.294 Million cell updates/sec

Tabular output not generated.

Title: >US-08-287-669-18
 Description: (2701-3700) from US08287669.seq (4 of 10)
 Perfect Score: 1000
 N.A. Sequence: 2701 TTGAGTGAATATATTTA.....GTAAAGAGCGATGATTT 3700
 Comp: AACTTCACCTTATATAAAT.....CATTTTCGCGTACGTAA

Scoring table: TABLE default
 Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database:

emb158
 1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
 7:em_in 8:em_ov 9:em_ov 10:em_ov 11:em_pat 12:em_ph
 13:em_pl 14:em_ro 15:em_sts 16:em_vi
 genbank111

Database:
 17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
 22:gb_in2 23:gb_ov 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
 28:gb_pl2 29:gb_pl1 30:gb_pl2 31:gb_pl3 32:gb_ro
 33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 11.371; Variance 8.664; scale 1.312

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1000	100.0	7653	21	CELCED3A	Caenorhabditis elegans	0.00e+00
2	930	93.0	39908	21	CEC48D1	Caenorhabditis elegans	0.00e+00
3	930	93.0	314495	19	CEY67H2	Caenorhabditis elegans	0.00e+00
4	134	13.4	27236	21	CEY32F6A	Caenorhabditis elegans	1.51e-50
5	134	13.4	36879	19	CEY07H11	Caenorhabditis elegans	1.51e-50
6	134	13.4	187816	19	CEY32F6	Caenorhabditis elegans	1.51e-50
7	132	13.2	36454	19	AC006670	Caenorhabditis elegans	1.70e-49
8	132	13.2	150963	19	AC006727	Caenorhabditis elegans	1.70e-49
9	132	13.2	274626	19	AC006903	Caenorhabditis elegans	1.70e-49
10	131	13.1	30522	21	CELM01H9	Caenorhabditis elegans	5.67e-49
11	131	13.1	274626	19	AC006903	Caenorhabditis elegans	5.67e-49
12	131	13.1	340801	19	AC006751	Caenorhabditis elegans	5.67e-49
13	124	12.4	38112	22	CELR08C7	Caenorhabditis elegans	2.59e-45

RESULT	1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS	1	CELCED3A	7653 bp <td>DNA <td>INV <td>23-FEB-1994</td> </td></td>	DNA <td>INV <td>23-FEB-1994</td> </td>	INV <td>23-FEB-1994</td>	23-FEB-1994
DEFINITION	1	Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
ACCESSION	129052					
VERSION	9456416					
KEYWORDS	129052.1	GI:456416				
SOURCE	1	cell death protein; interleukin-1 beta converting enzyme.				
ORGANISM	1	Caenorhabditis elegans (strain N2) DNA.				
REFERENCE	1	Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.				
AUTHORS	1	(bases 1 to 7653)				
TITLE	1	Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.				
JOURNAL	1	The C. elegans cell death gene ced-3 encodes a protein similar to				
MEDLINE	1	mammalian interleukin-1b-converting enzyme				
FEATURES	1	Cell 75, 641-652 (1993)				
source	1	94061982				
location/Qualifiers	1	1..7653				
/organism="Caenorhabditis elegans"	1					
/strain="N2"	1					
/db_xref="taxon:6239"	1					
/gene="ced-3"	1					
/number="1"	1					
/join(2232..2366,2430..2575,2854..3107,4303..4634,	1					
5547..5760,5815..5942,6298..6537,7013..7075)	1					
/gene="ced-3"	1					
/join(2232..2366,2430..2575,2854..3107,4303..4634,	1					
5547..5760,5815..5942,6298..6537,7013..7075)	1					
/gene="ced-3"	1					
/note="codes for a protein similar to mammalian	1					
interleukin-1b-converting enzyme"	1					
/codon_start=1	1					
/product="cell death protein"	1					

ALIGNMENTS

LOCUS A0191954 431 bp DNA GSS 03-NOV-1998
 DEFINITION HS-3228_B2_F08_MR_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3228 COL=16 Row=L, genomic survey sequence.

ACCESSION A0191954
 NID 93590576
 VERSION A0191954.1 GI:3590576
 KEYWORDS GSS.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 431)
 MAHAIRAS,G.G., WALLACE,J.C., SMITH,K., SWARTZELL,S., FURLONG,J., SHAKER,R., SCHMIDT,S., TRACOFF,R. and HOOD,L.E.
 Construction of a Characterized Clone Resource for Genomic Sequencing
 JOURNAL Unpublished (1998)

COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3228 row: L column: 16
 Class: BAC ends
 High quality sequence stop: 431.

FEATURES
 source
 Location/Qualifiers
 1..431
 /organism="Homo sapiens"
 /note="Organ: sperm; Vector: pBelOAC11; BAC Clones in E-Coli DH10B"
 /db_xref="taxon:9606"
 /clone_plate="3228 COL=16 Row=L"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"

BASE COUNT 186 a 70 c 59 g 116 t
 ORIGIN

Query Match 2.6%; Score 26; DB 40; Length 431;
 Best Local Similarity 86.1%; Pred. No. 1.96e-04;
 Matches 31; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 48 GAAAAAATCAAAACTATGCAAAATACAAAAAT 83
 ||||||||||||||| ||| |||| |||||
 QY 4092 GAAAAAATCAAAATTCGCGAAATTCAAAAAT 4127

Search completed: Sat Aug 7 19:20:22 1999
 Job time : 1739 secs.

BASE COUNT	49 a	33 c	38 g	74 t			
ORIGIN	/lab_host="DH10B"						
Query Match	2.6%	Score 26;	DB 26;	Length 194;			
Best Local Similarity	93.3%	Pred. No. 1.96e-04;					
Matches	28;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;			
Db	24	TTAAATCCATTTTATTTATTTTCACAT	53				
Oy	3902	TTAAATACATTTTGTATTTTCAACAT	3931				
RESULT	14						
LOCUS	AA971176	260 bp	mRNA	EST			
DEFINITION	op11.fil.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1575405 3',			20-MAY-1998			
ACCESSION	AA971176						
NID	9316466						
VERSION	AA971176.1	GI:3146466					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;						
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.						
TITLE	1 (bases 1 to 260)						
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.						
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
	Tumor Gene Index						
	Unpublished (1997)						
	On Jan 19, 1998 this sequence version replaced gi:2285245.						
	Contact: Robert Strausberg, Ph.D.						
	Tel: (301) 496-1550						
	Email: Robert.Strausberg@nih.gov						
	Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.						
	Emmert-Buck, M.D., Ph.D.						
	cDNA Library Preparation: Stratagene, Inc.						
	cDNA Library Arrayed by: Greg Lennon, Ph.D.						
	DNA Sequencing by: Washington University Genome Sequencing Center						
	Clone distribution: NCI-CGAP clone distribution information can be						
	found through the I.M.A.G.E. Consortium/LNL at:						
	www.bio.lnl.gov/bbrp/Image/Image.html						
FEATURES	Seq primer: -40ml3 fwd. ET from Amersham						
source	High quality sequence stop: 259.						
	Location/Qualifiers						
	1..260						
	/organism="Homo sapiens"						
	/note="Organ: kidney; Vector: Bluescript SK-; Site_1:						
	EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:						
	Oligo dt. Pooled kidney tumors. 5' adaptor sequence: 5'						
	GAATCGACGACGAG 3' 3' adaptor sequence: 5'						
	CTGAGCTTTTATTTTATTTT 3' Average insert size: 1.0 kb."						
	/db_xref="taon.9606"						
	/clone="IMAGE:1575405"						
	/clone_id="NCI_CGAP_Kid6"						
	/sex="mixed"						
	/tissue_type="kidney tumor"						
	/lab_host="SOIR (kanamycin resistant)"						
BASE COUNT	74 a	32 c	38 g	112 t			
ORIGIN	4 others						
Query Match	2.6%	Score 26;	DB 21;	Length 260;			
Best Local Similarity	72.2%	Pred. No. 1.96e-04;					
Matches	39;	Conservative	0;	Mismatches 15; Indels 0; Gaps 0;			
Db	200	TTCTAATTTTCTTGATATNTTTGGGCTTCCAAAGTTTGCGAAT	253				
Cp	4420	TTCAATCCTCTTCATGGAATATGATTGAGTGCTCCAGAACTTGCGAAT	4367				
RESULT	15						

TITLE	An Aspergillus nidulans EST Database
JOURNAL	Unpublished (1998)
COMMENT	On Aug 21, 1998 this sequence version

Contact: Bruce A. Roe, University of Oklahoma, broee@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel.: 405 325 4912

We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center
Seq primer: T3
High quality sequence start: 134

FEATURES	Location/Qualifiers
Source	1 245

```
/organism="Emericella nidulans"  
/strain="FGSC A26"
```

```

/notes=Vector: pBluescript SK-; Site_1:EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"
/db_xref="taxon:5072"

```

```

/collection_id=Aspergillus nidulans 24hr asexual
developmental and vegetative cdna lambda zap library
/tissue_type="vegetative mycelia, asexual structures"
58 2 40 3 91 4

```

BASE COUNT	67 a	58 c	40 g	81 t
ORIGIN				

Query Match	2.7%	Score 27	DB 24	Length 246
Best Local Similarity	80.0%	Pred. No. 1.80e-05		
Matches	36	Conservative	0	Mismatches 9
			Indels	0
			Gaps	0

D6	16	TTTTTTTTTTTTTTTACACACAAAAATGACCCCATGATTTTTT	60
		- - -	
QY	3912	TTTTTTGATTTTTTCAACATTCACATGATTAAACCCCATTAATTTTTT	3956

RESULT	9				
LOCUS	N98032	345 bp	mRNA	EST	18-NOV-1996
DEFINITION	2086C3 czapFPbd2.1, Debopam Chakrabarti		Plasmodium falciparum		
DESCRIPTION	clone PF2086C, mRNA sequence.				
SEQUENCE	100000				

```

VERSION      92072806
KEYWORDS     N98032.1  GI:1675066
SOURCE       malaria parasite P. falciparum

```

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
1 (bases 1 to 345)

REFERENCE	1 (bases 1 to 345)
AUTHORS	Dame, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.

TITLE Current status of the *Plasmodium falciparum* genome project
JOURNAL Mol. Biochem. Parasitol. 79, 1-12 (1996)
MEDLINE 97001675

COMMENT On Apr 14, 1993 this sequence version replaced gi:716382

Contact: Debopam Chakrabarti
Department of Molecular Biology and Microbiology
University of Central Florida
Orlando, FL 32816-3260

Tel: 407 384 2061
Fax: 407 384 3095

Email: dchak@pegasus.cc.ucf.edu
Seq primer: T3.

FEATURES	
Source	Location/Qualifiers
	1 345

```
/organism="Plasmodium falciparum"  
/strain="Dd2"
```

```

/notes=Vector: Lambda ZAP II; Site:1: EcoR I; Site:2: Xho I;
PolyA+ RNA, from asynchronous blood stage parasites of
the Dd2 isolate cultured in vitro, was reverse transcribed
using an oligo dt-Xho I primer. Second strand was
prepared using RNase H and DNA polymerase I. EcoR I
adapters were ligated to the cDNA, and it was digested
with Xho I. Prepared fragments were ligated into EcoR I +
Xho I digested lambda ZAP II vector.
/db_xref="taxon:5633"
/clone="PF2086C"
/clone_lib="czapppd2.1, Debopam Chakrabarti"
/lab_host="E. coli XL-1 blue"
BASE COUNT      137 a      32 c      16 g      158 t      2 others
ORIGIN

```

```
Query Match      2.7%; Score 27; DB 33; Length 345;
Best Local Similarity 64.4%; Pred. No. 1.80e-05;
Matches 56; Conservative 0; Mismatches 31; Indels 0; Gaps 0
```

DD 9/ AATGGAATGAAAAAATTAATTTTATTCATTNNACATNATTTGTATATAA 15
 || ||||| ||||| || ||||| || || ||||| || ||
 QY 3616 AATGGAATGACGAATAATAGCAATTTTATATAATTTTCCCTATTCATGTTGTGCAGAA 36

DB 157 TATAAATTAAATTTCTATACATTTTT 183
 | | | | | | | | | |
 QY 3676 AAATAGTAAAAAAGCGCATGATTTTT 3702

RESULT 10

DEFINITION
qds56s10.x1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:1733514
3' similar to contains Alu repetitive element; contains element XBR
repetitive element ;, mRNA sequence.
ACCESSION
AF262740

VERSION	AI262740.1	GI:3870943
KEYWORDS	EST.	
SOURCE	human.	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 431)

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On APR 14, 1993 this sequence version replaced a1:682609

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrr/image/image.html

Insert Length: 555 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 398.

FEATURES	Location/Qualifiers
source	1. .431

/organism="Homo sapiens"
Vector: pT7T3D-Pac (Pharmacia) with a modified
linker: Site 1: Not I. Site 2: Eco RI. 1st strand cDNA

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

Db 1611 GMTKTTTMMWMSMTW 1627
 QY 3786 TCTGCTCTTCAGATG 3802

RESULT 6
 LOCUS AF034173 2275 bp mRNA EST 30-MAR-1998
 DEFINITION AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
 clone ntcon2 contig, mRNA sequence.
 ACCESSION AF034173
 NID 92707735
 VERSION AF034173.1 GI:2707735
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 2275)
 AUTHORS Tripodis, N. and Ragoussis, J.
 TITLE Generation of a transcription map in the region immediately centromeric to human MHC across the 6p21.2-6p21.3 chromosomal boundary
 JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2045115.

BASE COUNT 438 a 619 c 470 g 599 t 149 others

ORIGIN

Query Match 3.5%; Score 35; DB 20; Length 2275;
 Best Local Similarity 13.8%; Pred. No. 1.39e-14;
 Matches 26; Conservative 93; Mismatches 66; Indels 3; Gaps 2;

Db 1484 WKRRGRKRRKMTGKMYKMYKMMAMCAMWYKMKMGKMKYKRYKRYSTYK 1543
 Cp 4287 TTGAGTCATGTTTGTATACCAATTTCTCTTTTGTGCGATTTTATTT 4228
 Db 1544 SMSWMTYTTTTCWCCTISMSKASACAMKMGYMSRS--SRSTWGYGSMSCYGMT 1601
 Cp 4227 CGCATTTTCATTCATTCATTCCTAGAAATTTGATTTCCCTTTTATTTTACAA 4168
 Cp 1602 KRY-YRYSMTGWTMTTMMWMSMTTMTTMTTMTTMTTMMWMTTCTMTKRRGA 1660
 Cp 4167 TTCTATGATGATGAAAAATGCAAAATGTATTTTATTCATTTTTCATTTTCGAGA 4108
 Db 1661 GTTTTGA 1668
 Cp 4107 ATTTTGA 4100

RESULT 7
 LOCUS R03411 352 bp mRNA EST 31-MAR-1995
 DEFINITION pk06002.s1 Kuwabara Mixed stage C. briggsae Caenorhabditis briggsae
 cDNA, mRNA sequence.
 ACCESSION R03411
 NID 6753147
 VERSION R03411.1 GI:753147
 KEYWORDS EST.
 SOURCE Caenorhabditis briggsae.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

REFERENCE
 AUTHORS
 Rhabditia; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 352)
 Hillier, L., Chappell, B., Chisoe, S., Clark, N., Couch, J.,
 Dubuque, T., Hawkins, M., Holman, M., Hultman, M., Kucaba, T.,
 Kuwabara, P., Le, M., Mardis, E., Marra, M., Parsons, J., Rifkin, L.,
 Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Wohlmann, P. and
 Wilson, R.
 Washington University Caenorhabditis briggsae EST project
 Unpublished (1995)
 Other ESTs: pk06002.r1
 Contact: Marra MA
 Washington University Genome Sequencing Center
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1455
 Fax: 314 286 1810
 Email: mmara@wustl.edu
 PCR_F: TGTAAACGACGCGATGAGTACGAGTTCAGCCTG
 PCR_B: CAGGAACGATGATGACCTTATGATTTCTTCAGGGA
 Source: Washington University Genome Sequencing Center
 PCR amplified DNA is available from Washington University Genome
 Sequencing Center. Aliquots of the library may be requested from P.
 Kuwabara (pek@wustl.edu, cam.ac.uk).
 Insert length: 839 Std Error: 0.00
 Seq primer: Commercially available -21M13 dye primer
 High quality sequence stop: 359.

FEATURES
 source
 Location/Qualifiers
 1..352
 /organism="Caenorhabditis briggsae"
 /strain="G16 Gujarat"
 /note="Vector: lambda gt10; Site 1: EcoRI; Site 2: EcoRI;
 Stage: mixed, sex: hermaphrodite. Library construction:
 First strand oligo(dT) primed. Second strand was as per
 Gubler/Hofman. Ligated to EcoRI adaptors. Library is
 non-directional. Library is non-normalized. Library is
 constructed by P.E. Kuwabara. Additional details on
 construction of the library are described in P.E.
 Kuwabara and S. Shan, NAR 22: 4414 - 4418 (1994). Adaptor
 sequence: GAATTC CGTGGCTGCG
 /db_xref="taxon:6238"
 /clone_lib="Kuwabara Mixed stage C. briggsae"

BASE COUNT 96 a 69 c 90 g 87 t 10 others

ORIGIN

Query Match 3.2%; Score 32; DB 19; Length 352;
 Best Local Similarity 72.9%; Pred. No. 5.13e-11;
 Matches 62; Conservative 0; Mismatches 21; Indels 2; Gaps 2;

Db 1 GAACATTCGCGAGATCCACTAGAAAGCGGCGAAGCGGANTTAAGACACACTAA 60
 QY 4518 GACACCTTGAGCAGATGCCAACAGGAGATGTATCAAGAGCGGAC-AAAGACAAT-CTTA 4575
 Db 61 CGAATTTGTCAGATGATGGGATA 85
 QY 4576 CCAATTTGTCAGATGATGGGATA 4600

RESULT 8
 LOCUS A1212009 246 bp mRNA EST 19-OCT-1998
 DEFINITION w80101.r1 Aspergillus nidulans 24hr asexual developmental and
 vegetative cDNA lambda zap library Emericella nidulans cDNA clone
 w80101.5', mRNA sequence.
 ACCESSION A1212009
 NID 93773951
 VERSION A1212009.1 GI:3773951
 KEYWORDS EST.
 SOURCE Emericella nidulans.
 ORGANISM Emericella nidulans.
 Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
 Eurotiales; Trichocomaceae; Emericella.
 REFERENCE 1 (bases 1 to 246)
 Kuper, D., Gray, J., Hausner, J., Lal, H., Martin, W., Aramayo, R.,
 Prade, R. and Roe, B.

```

/organism="Oryza sativa"
/cultivar="Milyang23"
/notes="Vector: pluscript SK(+): Site.1: EcoRI; Site.2:
XhoI: Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/DB_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA library"

```

	Matches	17; Conservative	75; Mismatches	43; Indels	2; Gaps	2;
D _b	1491	RKRRTGATKRRYRMMAMMAMCMAACWYYIKMKRMCKMYIRKKYTTSTIYYSSMSMWYM	1550			
O _y	3668	GTCGAGAAAAATACTAATAAAAAGCCCATGCACTTTTTCGCACAT-TTTTTAACATCGAACACA	3726			
D _b	1551	YTTYTYVYWCQCTSMKSASCAMNRMWMGYMSRSRSRYWGYSNGCYGTAKRYRYSWT	1610			
X _y	3727	GCTCACCTCATCATGCTTGATMACGAGNACCGG-AGAAATAACACACATCTTTCTGGTC	3785			

[illegible]

```

/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_1lb="Rice Immature Seed Lambda ZAP11 cDNA library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT      5 a 21 c 12 g 35 t 179 others
ORIGIN
Query Match      5.38; Score 53; DB 17; Length 252;
Best Local Similarity 11.5%; Pred. No. 2,16e-38;
Matches 25; Conservative 110; Mismatches 80; Indels 3; Gaps 2;

Db 19 MTTSYBCHGHBVWVCYASHGNYMVSVCNCTBCTHDDCKVMNNSITWGTVMVBNVSGDMH 78
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Cp 4571 ATTGTCCTTGTCGGCCTTGTCACATTCCTCGGTGGCATCTGCTCA--ACGTTCATTA 4514
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ||||

Db 79 YWBVBNTRKVDVGNHTRCSRWRBVTIRAHNHDYTNGBVYNNNDYHMHMBMYBRTGCMTC 138
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ||||

Cp 4513 TTTATGATGAGGCACATCCACAGAGACTCGAGAGATTCTGTACATGCTTTCTCGCG 4454
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ||||

Db 139 MTCBHNNTKCTASGWHSTNNDYKSCSTNTWGYTBSYDKSMHGVCSDBVKYHTKVSFTFR 198
    ::::: ||::: ||::: ::::: ::::: ::::: ::::: ::::: ::::: ||||
Cp 4453 AAAACACGGGCTTATGGTGTGTCATGACAAAGATTCATCTCTTCATGAAATATGAT 4394
    ::::: ||::: ||::: ::::: ::::: ::::: ::::: ::::: ::::: ||||

Db 199 ATRSYTCVRKYCVMM-MTKKKVKKYHVBVBBGHBHDSK 235
    ::::: ||::: ||::: ::::: ::::: ::::: ::::: ::::: ::::: ||||
Cp 4393 TGAAGTGTGCACAGAGCTTGCTGGAATGAGCAATTACG 4356
    ::::: ||::: ||::: ::::: ::::: ::::: ::::: ::::: ::::: ||||

```

FEATURES	source
RESULT	3
LOCUS	AA754458 247 bp mRNA
DEFINITION	97SN1784 Rice Immature Seed Lambda ZAPI cDNA Library <i>Oryza sativa</i>
ACCESSION	CDNA clone 97SN1784, mRNA sequence.
NID	AA734458
VERSION	92801164
KEYWORDS	AA754458.1 GI:2801164
SOURCE	EST.
ORGANISM	<i>Oryza sativa</i> .
REFERENCE	<i>Oryza sativa</i> . Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza. 1 (bases 1 to 247) Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,M.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y. Large-scale Sequencing Analysis of ESTs from Rice Immature Seed unpublished (1998) On Jan 14, 1998 this sequence version replaced gi:1797455.
TITLE	Contact: Eun M.Y.
JOURNAL	Department of Cytogenetics
COMMENT	National Inst. of Agri. Sci. and Tech, RDA Suwon, Kyunggi-do, Korea Tel: 82 331 290 0301 Fax: 82 331 290 0307 Email: myeun@sun20.asti.re.kr Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhna@mbioserver.myongji.ac.kr Seq primer: M3 reverse primer. Location/Qualifiers 1..247

```

Location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/acc="F01493"
/notes="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/submitter="taxon:4530"
/imap="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed lambda ZAPII cDNA library"

```

(TM)

ford Molecular Ltd

ing Smith-Waterman algorithm

1361 034 Million updates/sec
maspat time 1/21.60 seconds

seq (5 of 10)

.....AACAGTCTACGTACCCGAT

ises x 2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

gb_gss5 42:gb_gss6

03; scale 4.102

al score distribution.

Description	Prod. No.
97SN1787 Rice Immature	8.426-463
97SN1787 Rice Immature	2.156-384
97SN1784 Rice Immature	4.576-228
97SN1784 Rice Immature	4.846-111
AF034173 Human mRNA (T	4.846-111
AF034173 Human mRNA (T	1.396-131
PK060231 Kiwibara Mt	5.136-111
w8801a1.t1 aspergillus	1.806-070
2008C6 czapADp2.1, De	1.806-070
97SN01.x1 Soares Test	1.806-070

C	11	27	2.7	490.37	B88472	C	11	27	2.7	490.37	B88472	C	11	27	2.7	490.37	B88472	C	11	27	2.7	490.37	B88472	
	12	27	2.7	558.31	H23583		12	27	2.7	558.31	H23583		12	27	2.7	558.31	H23583		12	27	2.7	558.31	H23583	
	13	26	2.6	194.26	A4149399		13	26	2.6	194.26	A4149399		13	26	2.6	194.26	A4149399		13	26	2.6	194.26	A4149399	
	14	26	2.6	431.20	AA971176		14	26	2.6	431.20	AA971176		14	26	2.6	431.20	AA971176		14	26	2.6	431.20	AA971176	
	15	26	2.6	460.20	AA191954		15	26	2.6	460.20	AA191954		15	26	2.6	460.20	AA191954		15	26	2.6	460.20	AA191954	
	16	26	2.6	436.39	AA224642		16	26	2.6	436.39	AA224642		16	26	2.6	436.39	AA224642		16	26	2.6	436.39	AA224642	
	17	26	2.6	499.37	B32672		17	26	2.6	499.37	B32672		17	26	2.6	499.37	B32672		17	26	2.6	499.37	B32672	
	18	26	2.6	514.41	AA2863415		18	26	2.6	514.41	AA2863415		18	26	2.6	514.41	AA2863415		18	26	2.6	514.41	AA2863415	
	19	25	2.5	385.40	AA0166653		19	25	2.5	385.40	AA0166653		19	25	2.5	385.40	AA0166653		19	25	2.5	385.40	AA0166653	
	20	25	2.5	388.39	AA0210219		20	25	2.5	388.39	AA0210219		20	25	2.5	388.39	AA0210219		20	25	2.5	388.39	AA0210219	
	21	25	2.5	421.38	AA064496		21	25	2.5	421.38	AA064496		21	25	2.5	421.38	AA064496		21	25	2.5	421.38	AA064496	
	22	25	2.5	426.28	AA1561183		22	25	2.5	426.28	AA1561183		22	25	2.5	426.28	AA1561183		22	25	2.5	426.28	AA1561183	
	23	25	2.5	436.24	AA1267334		23	25	2.5	436.24	AA1267334		23	25	2.5	436.24	AA1267334		23	25	2.5	436.24	AA1267334	
	24	25	2.5	436.38	AA0245342		24	25	2.5	436.38	AA0245342		24	25	2.5	436.38	AA0245342		24	25	2.5	436.38	AA0245342	
	25	2.5	447.40	AA040053		25	2.5	447.40	AA040053		25	2.5	447.40	AA040053		25	2.5	447.40	AA040053		25	2.5	447.40	AA040053
	26	2.5	447.16	AA0436494		26	2.5	447.16	AA0436494		26	2.5	447.16	AA0436494		26	2.5	447.16	AA0436494		26	2.5	447.16	AA0436494
	27	2.5	450.28	AA1502086		27	2.5	450.28	AA1502086		27	2.5	450.28	AA1502086		27	2.5	450.28	AA1502086		27	2.5	450.28	AA1502086
	28	2.5	474.42	AA242689		28	2.5	474.42	AA242689		28	2.5	474.42	AA242689		28	2.5	474.42	AA242689		28	2.5	474.42	AA242689
	29	2.5	478.40	AA067917		29	2.5	478.40	AA067917		29	2.5	478.40	AA067917		29	2.5	478.40	AA067917		29	2.5	478.40	AA067917
	30	2.5	499.39	AA121161		30	2.5	499.39	AA121161		30	2.5	499.39	AA121161		30	2.5	499.39	AA121161		30	2.5</		

ALIGNMENTS

[illegible]

Contact: Eun M.Y.
 Department of Cytogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyungido, Korea
 Tel.: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.asri.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
 University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
 Seq primer: M13 Reverse Primer.
 location/Qualifiers
 1..252
 /organism="Oryza sativa"
 source

```

FT      Intron      6538..7012      /*tag= x
FT      /number= Intron_7
FT      repeat_unit  6567..6625
FT      /*tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit  6905..6965
FT      /*tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon        7012..7075
FT      /*tag= aa
FT      /number= Exon_8
PN      WO9325685-A.
PD      23-DEC-1993.
PF      14-JUN-1993; 005701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
DR      WPI; 94-007542/01.
DR      P-PSDB; R53282.
PT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
PS      develop agents to increase or prevent cell death in organisms
PS      Claim 14; Fig 4; 127p; English.
CC      The sequences given in Q64735-45 represent mutations of the C. elegans
CC      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
SQ      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

```

Query Match 12.7%; Score 127; DB 9; Length 7653;

Best Local Similarity 78.0%; Pred. No. 2.76e-56;

Matches 309; Conservative 0; Mismatches 68; Indels 19; Gaps 13;

```

Db      3463 gttcgcaagat-ttttgatlttttcattcaagatatgcttataacacataatat 3521
          ||||||| || ||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp      4087 GTTTCGACATATTTTGGCTTGACTTCATTCAGATATGCTTATTACATCATTAATTAT 4028
          || ||| ||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      3522 cattaatgtaatttctttagaatttgggttcttcttagtatagtcttaactttg 3581
          || ||| ||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp      4027 CATTAAAGGTAATATCTTGAGAAATTTGG--TTTTCGCTCAATATCTCTACTTTT 3970
          || ||| ||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      3582 aaattgcacaagaaataatcattgtgttgcata-tatgaatgaaccaaaatagcaat- 3639
          || ||||||| ||||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
Cp      3969 AAGTGTCTCAGCAAAAATAATGAGGCTTAATCATGTGATGTTGAAAAATACAAAAATG 3910
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      3640 t-attta-tatatltccc-tattcattgtgtgcagaanaatagtaaaaaa-gc--gca 3693
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp      3909 TATTTAATACATTTTCCCTATTCAT-TTGGCGAGAAAGCTGTAATAAAACGATGCA 3851
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      3694 tg-----catttgcacattttt-acatggaagacagctcacttcacatgctgaaga 3747
          || ||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp      3850 TTTTTCATATATTCGACATTTTTTTACATCGACGAGATCCCATTTTCACATGCTGAAGA 3791
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

Db      3748 cgaagacgcggaagaataacacacacatttctgcgtctctcgtcttcagcatgtgaat 3807
          ||||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Cp      3790 CGAGAGACGCAGAAAAGATGTGTGTAATTTCTCCGGTCTCTCGCTTCAGCATGTGAAGT 3731
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      3808 gggatctcgtcagatgtaaaaaaatgctgaataatg 3843
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp      3730 GAGCTGTGTTGATGTAAAAA-TGTCGAAAAAATG 3696
          || ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

Search completed: Sat Aug 7 19:24:43 1999
Job time : 241 secs.

Query Match	99.8%	Score 998	DB 9	Length 7653
Best Local Similarity	99.9%	Pred. No. 0.00e+00		
Matches	999	Conservative	0	Mismatches 1; Indels 0; Gaps
Db	3601	tcattggtttgttcataatgaatgacaaataagcaatttttatatatatttccctat	3660	
Qy	3601	tcattggtttgttttcataatgaatgacaaataagcaatttttatatatatttccctat	3660	
Db	3661	tcattgtgcaagaataatgaataaagcgatgcattttcgcacatttttataatga	3720	
Qy	3661	tcattgttgcagaaaaatagtaaaaaagcgcatgtatttttgcacatttttataatga	3720	
Db	3721	acgaagcttcattcacatgctgaagacggaacgcggaataataccacatcttct	3780	
Qy	3721	acgaagcttcattcacatgctgaagacggaacgcggaataataccacatcttct	3780	
Db	3781	gggtcttcgcttcacagcatgtgaatgggactcgtgcgatgtaaaaaatgctgaata	3840	
Qy	3781	gggtcttcgcttcacagcatgtgaatgggactcgtgcgatgtaaaaaatgctgaata	3840	
Db	3841	atgtaaaaaatgcatgsggtttttttacatttctgcacaaatgaatgagggaaaaatgt	3900	
Qy	3841	atgtaaaaaatgcatgsggtttttttacatttctgcacaaatgaatgagggaaaaatgt	3900	
Db	3901	attaaatacaatttttgtatatttcaacatcacatgatacccatatbttttgt	3960	
Qy	3901	attaaatacaatttttgtatatttcaacatcacatgatacccatatbttttgt	3960	
Db	3961	gagcaacttaaaagtgaataatattagagcggaacccaattcttcaagatatcc	4020	
Qy	3961	gagcaacttaaaagtgaataatattagagcggaacccaattcttcaagatatcc	4020	
Db	4021	ttattgtlaattatagatgttaataagcatcttggaaatgaagtcgaacaaatagt	4080	
Qy	4021	ttattgtlaattatagatgttaataagcatcttggaaatgaagtcgaacaaatagt	4080	
Db	4081	gcgaaacactgtaaaaaatcaaaaatctctcgaaaattgtaaaaaatgcattaataata	4140	
Qy	4081	gcgaaacactgtaaaaaatcaaaaatctctcgaaaattgtaaaaaatgcattaataata	4140	
Db	4141	tttttgcattttttacaatagaaatgtagaataatgaaggaatacaaatcttca	4200	
Qy	4141	tttttgcattttttacaatagaaatgtagaataatgaaggaatacaaatcttca	4200	
Db	4201	gaggatlaattgtatgaacatttggaaatzaaaatgtgycgaacgctcaaaaaagaga	4260	
Qy	4201	gaggatlaattgtatgaacatttggaaatzaaaatgtgycgaacgctcaaaaaagaga	4260	
Db	4261	aatttgggtatcaaatcgatccttaaaccaacatlttcagcatcgcgcaacttcat	4320	
Qy	4261	aatttgggtatcaaatcgatccttaaaccaacatlttcagcatcgcgcaacttcat	4320	
Db	4321	tcacggagatcctctctctcgagataagtcgaagtcgaatgcgcatttggagaagtt	4380	
Qy	4321	tcacggagatcctctctctcgagataagtcgaagtcgaatgcgcatttggagaagtt	4380	
Db	4381	cttgaaccaactcaatacatatltccatgaaagagatatgaacttltcgatgcaccaaca	4440	
Qy	4381	cttgaaccaactcaatacatatltccatgaaagagatatgaacttltcgatgcaccaaca	4440	
Db	4441	taagcgtgttttcgacgagaaaaacatgtacagaaacttctcgagctccctgtggaatgt	4500	
Qy	4441	taagcgtgttttcgacgagaaaaacatgtacagaaacttctcgagctccctgtggaatgt	4500	
Db	4501	ggcccatcataaataatgaacacttggagcagatgcgaacgagaaatgtataccaaagcg	4560	
Qy	4501	ggcccatcataaataatgaacacttggagcagatgcgaacgagaaatgtataccaaagcg	4560	
Db	4561	acaagacaactcttaccgaatttgttcaatgcatgagcta 4600		
Qy	4561	acaagacaactcttaccgaatttgttcaatgcatgagcta 4600		

CC splicing mutations (see also Q64735-45). These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 Sequence 7653 BP; 2429 A; 1453 C; 1272 G; 2499 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;
 Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3601 tcatgtgtgttgcataatgacacaaatgacacatttttatataatcccccat 3660
 |||||||
 Qy 3601 TCATGTGTTGTCATATGAATGAGAAAATACCAATTTTATATTTTCCCTAT 3660
 |||||||
 Db 3661 tcatgtgtgcagaaaatagtaaaaagcgatcatcttttcgacatttttacatga 3720
 |||||||
 Qy 3661 TCATGTGTGCGAGAAAATAGTAAAAAGCGATGCAATTTTCGACATTTTACATGCA 3720
 |||||||
 Db 3721 acgacagctcacttcacatgctgaagacgagagacgagagaaatacacacatcttct 3780
 |||||||
 Qy 3721 ACGACAGCTCCTCATCTCAGTGAAGACGAGACGGGAGAAATACGACATCTTCTCT 3780
 |||||||
 Db 3781 ggcgtctctgcttcagcatgtgaatggatcctgctgctgataaaaaatgtcgata 3840
 |||||||
 Qy 3781 GCGTCTCTGCTCTGCGATGTGAATGGATCTCGTGTGATGTAATAAATGTGCAATA 3840
 |||||||
 Db 3841 atgtaaaaatgcacgcttttttttaccatttcgcacaaatgataaggggaaaaatgt 3900
 |||||||
 Qy 3841 ATGTAAAAATGCATCGCTTTTTCACATTTTTCGACAAATGATAGGGGAAAATCT 3900
 |||||||
 Db 3901 attaaaatacatcttttgatatttcaacatcacatgataaaccatatttttcgt 3960
 |||||||
 Qy 3901 ATTAATAATCATTTTGTATTTTTCACATCAGATTAACCCATATTTTTCGAT 3960
 |||||||
 Db 3961 gagcaacttaaaaaagtaagataatagagcgaaaaaccaaatttcttaagatatcc 4020
 |||||||
 Qy 3961 GAGCAACTTAAAAAGTAGAATATTAGAGCGAAAACCAAAATTTCTTCAAGATATTAC 4020
 |||||||
 Db 4021 ttatattgataatataagattgaataagcatatcttgataagaaagcagcaaaatagt 4080
 |||||||
 Qy 4021 TTTATTGATATATTATGATGTTATATAGCATATCTTGATGGAAGTCACCAAAATATCT 4080
 |||||||
 Db 4081 gcgaaacacctgaaaaaaatcaaaaatctgcgaaaaattgaaaaaaatgcatataataca 4140
 |||||||
 Qy 4081 GCGAAACACCTGAAAAAATCAAAAATCTCGGAAAATGAAAAAATGCATTAATAACA 4140
 |||||||
 Db 4141 tttttgcatcttttcatacatcagatgataagataaattgaaggaatcaaatcttca 4200
 |||||||
 Qy 4141 TTTTTCATTTTCTCATCTCATCTGATGATGTAGAAAATTTAAAAGGGAATCAAAAATTTCTCA 4200
 |||||||
 Db 4201 gaagataataatgaatgaacacattgcgaataataatgctgcgaaaagtaaaaaagaga 4260
 |||||||
 Qy 4201 GAGGATATATATGATGAACAATGGGAATTTAAATGTGCCGAACGTCAAAAAAGAGA 4260
 |||||||
 Db 4261 aatttggatatacaaatcgatcctaaacacacacatttcagatcgcgcaacttctcat 4320
 |||||||
 Qy 4261 AATTGGGATCAAAATCGATCTTAACCAACACATTTTCAGATCTGCCCAACTCTTCAT 4320
 |||||||
 Db 4321 tcacggatgctcttcttcggtatagttcaagtctgatatgctatttcagcaagctt 4380
 |||||||
 Qy 4321 TCACCGGATGCTCTCTCTCGATACAGTTCAAGTGTATGCTCTATTCAGCAAGCTT 4380
 |||||||
 Db 4381 ctggacaactcaatacatatcctcagaagagatatgaaactgtgcgtgcacacaaca 4440
 |||||||
 Qy 4381 CTGGACCAACTCAATATATATTTCCATGAAGAGATATGAACCTTGTGCGATGCACCAACA 4440
 |||||||
 Db 4441 taagccgtgttttcgacgagaaaaacatgtacagaaactctcagatctcgttggaatgt 4500
 |||||||
 Qy 4441 TAAGCCGTGTTTTCGACGAGAAAACCATGTACAGAAATCTCTGAGTCTCTCGGAAATGT 4500
 |||||||

Db 4501 gccctcataataatgaaactttgagcagatgccaacacggaatgttaccagggccg 4560
 |||||||
 Qy 4501 GCCCTCATATAATTAAGAACTTTGAGCGATGCGCAACCGAATGGTACCAAGCCG 4560
 |||||||
 Db 4561 acaagacaatcttacaatttgcagatgcatggctta 4600
 |||||||
 Qy 4561 ACAAGCAATCTTACCAATTTGTCAGATGATGGGCTA 4600
 |||||||

RESULT 13
 ID 064737 standard; DNA; 7653 BP.
 AC 064737;
 DT 23-JUN-1994 (first entry)

DE ced-3 (G5757A) gene.
 KW C. elegans; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.

PH Key location/Qualifiers
 FT repeat_unit 1356..1472

FT /*tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1490..1614"

FT repeat_unit 1490..1614
 FT /*tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356..1472"

FT exon 2232..2366
 FT /*tag= c
 FT /number= Exon_1
 FT 2367..2429

FT intron
 FT /*tag= d
 FT /number= Intron_1
 FT 2450..2575

FT exon
 FT /*tag= e
 FT /number= Exon_2
 FT 2576..2853

FT intron
 FT /*tag= f
 FT /number= Intron_2
 FT 2854..3107

FT exon
 FT /*tag= g
 FT /number= Exon_3
 FT 3108..4302

FT intron
 FT /*tag= h
 FT /number= Intron_3
 FT 3126..3243

FT repeat_unit
 FT /*tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329..3396"

FT repeat_unit
 FT /*tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782..4070"

FT repeat_unit
 FT /*tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487..3759"

FT repeat_unit
 FT /*tag= l
 FT /number= Exon_4
 FT 4635..5546

FT intron
 FT /*tag= n
 FT /number= Intron_4
 FT 4688..4719
 FT repeat_unit
 FT /*tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"

FT repeat_unit 5221..5330

```

FT      /tag= a
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t repeat at 1490-1614"
FT      1490..1614
FT      /tag= b
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 1356-1472"
FT      2232..2366
FT      /tag= c
FT      /number= Exon_1
FT      2310
FT      /tag= d
FT      /note= "C>T, from allele n1040"
FT      2367..2429
FT      /tag= e
FT      /number= Intron_1
FT      2430..2575
FT      /tag= f
FT      /number= Exon_2
FT      2487
FT      /tag= g
FT      /note= "G>A, fromm allele n718"
FT      2576..2853
FT      /tag= h
FT      /number= Intron_2
FT      2854..3107
FT      /tag= i
FT      /number= Exon_3
FT      3108..4302
FT      /tag= j
FT      /number= Intron_3
FT      3126..3243
FT      /tag= k
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT      3329..3396
FT      /tag= l
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT      3487..3759
FT      /tag= m
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT      3782..4070
FT      /tag= n
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT      4303..4634
FT      /tag= o
FT      /number= Exon_4
FT      4635..5546
FT      /tag= p
FT      /number= Intron_4
FT      4688..4719
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5221..5330
FT      /tag= r
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5547..5760
FT      /tag= s
FT      /number= Exon_5
FT      5757
FT      /tag= t
FT      /note= "G>A, from allele n2433"
FT      5761..5814
FT      /tag= u
FT      /number= Intron_5
FT      5815..5942
FT      /tag= v
FT      /number= Exon_6

```

```

FT      mutation
FT      5940
FT      /tag= w
FT      /note= "C>T, from allele n1165"
FT      5943..6297
FT      /tag= x
FT      /number= Intron_6
FT      6062..6138
FT      /tag= y
FT      /rpt_type= INVERTED
FT      6297
FT      /tag= z
FT      /note= "G>A, fromm allele n717"
FT      6298..6537
FT      /tag= aa
FT      /number= Exon_7
FT      6322
FT      /tag= ab
FT      /note= "C>T, from allele n1949"
FT      6372
FT      /tag= ac
FT      /note= "G>A, from allele n1286"
FT      6434
FT      /tag= ad
FT      /note= "C>T, fromm alleles n1129 and n1164"
FT      6485
FT      /tag= ae
FT      /note= "C>T, from allele n2430"
FT      6535
FT      /tag= af
FT      /note= "G>A, from allele n2426"
FT      6538..7012
FT      /tag= ag
FT      /number= Intron_7
FT      6567..6625
FT      /tag= ah
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      /tag= ai
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7013..7075
FT      /tag= aj
FT      /number= Exon_8
FT      7020
FT      /tag= ak
FT      /note= "C>T, from allele n1163"
FT      mutation
FT      MO9325685-A.
FT      23-DEC-1993.
FT      PD 14-JUN-1993; U05701.
FT      PR 12-JUN-1992; US-897788.
FT      PR 20-NOV-1992; US-979638.
FT      PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      PI Horvitz HR, Shaham S, Yuan J;
FT      DR WPI; 94-007542/01.
FT      DR P-PSDB: R47466.
FT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT      develop agents to increase or prevent cell death in organisms
FT      PS Claim 2; Fig 4; 127P; English.
FT      CC This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
FT      CC was identified as the ced-3 transcript and was most abundant in
FT      CC embryos, but was also detected in larvae and young adults, suggesting
FT      CC that ced-3 is expressed not only in cells undergoing cell death. The
FT      CC four largest introns as well as sequences 5' of the start codon
FT      CC contain repetitive elements, some of which have been characterised
FT      CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
FT      CC and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
FT      CC highly hydrophilic with no significant hydrophobic region that might
FT      CC be a transmembrane region. One region of Ced-3 is very rich in serine.
FT      CC It is thought that this region is involved in protein-protein inter-
FT      CC actions, similar to acid blobs in transcription factors. Of the
FT      CC mutations which occur within the ced-3 gene, eight of the mutations
FT      CC are missense mutations, two are nonsense mutations and two are putative

```

```

FT      intron      5943..6297
FT      /tag= t
FT      /number= Intron_6
FT      repeat_region 6062..6138
FT      /tag= u
FT      /rpl_type= INVERTED
FT      exon      6298..6537
FT      /tag= v
FT      /number= Exon_7
FT      mutation 6322
FT      /tag= w
FT      /note= "C>T, from allele n1949"
FT      intron 6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit 6567..6625
FT      /tag= y
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit 6905..6965
FT      /tag= z
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon 7012..7075
FT      /tag= aa
FT      /number= Exon_8
PN      WO9325685-A.
PD      23-DEC-1993.
PR      14-JUN-1993; 005701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
PI      WPI; 94-007542/01.
DR      P-PSDB; R33283.
PT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT      develop agents to increase or prevent cell death in organisms
PS      Claim 14; Fig 4; 127pp: English.
CC      The sequences given in 064735-45 represent mutations of the C. elegans
CC      ced-3 gene. A 2.8 Kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
SQ      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db      3721 acgacagctcacttcacatgctgaagacgagagcgaggaataaccacatcttct 3780
Db      |||||||
Qy      3721 ACGACAGCTCCTTACATCTCTAAGACGAGAGACCGGAGAAATCCACACATCTTTCT 3780
Db      3781 ggcgtctcgtcttcagcatgtaaaatggatcctggtcgtatgtaaaaaatgctgata 3840
Qy      3781 GCGTCTCTCGTCTTACACATGTAAGATGGATCTCGGTGATGTAAATAATGTCGATA 3840
Db      3841 atgtaaaaaatgcagtcgcttttttcaactcttcycacaaatgaatagggggaaaaatg 3900
Qy      3841 ATGTAAAAATGATGCGTGTGTTTTCACACTTCTTCACAAAATGAATAGGGGAAAAATGT 3900
Db      3901 attaaaaatattttgtatcttttcaacacacacagattaccatatttttcgctt 3960
Qy      3901 ATTAAAAATACATTTTGTATTTTTCACATCTCAGATGATTAACCCATTAATTTTTCGTT 3960
Db      3961 gagcaacttataaaatgtagaataatagagcgagaaaccaaatttttcaagatatacc 4020
Qy      3961 GAGCAACTTAAAAAGTAGAATATTAGACGCAAAACCAAAATTTCTTCAAGTATTATAC 4020
Db      4021 ttattgataatatagatgltataaagcatcttgaaatgaatgaagtcagcaaaaatagt 4080
Qy      4021 TTATTGATATATTATAGATGTATTATAGCATATCTTGAAATGAAGTCACCAAAAATATGT 4080
Db      4081 gcgaaacacctgaaaaaaatcaaaaattctcgaaaattgaaaaaaatgcatataata 4140
Qy      4081 GCGAAACACCTGAAAAAAATCCAAAATTCGCAAAATTTGAAAAATGATTAATAATACA 4140
Db      4141 ttattgattttctcatcacatgtaatgtagaataatgaaggaatcaaatctca 4200
Qy      4141 TTTTGCATTTTCTCATCATCATGATGATGAGAAATTTAAAGGAAATCAAAATTTCTCA 4200
Db      4201 gaggataataatgtagaataatgtagaataatgaatgtagaataatgtagaataatgtaga 4260
Qy      4201 GAGGATATATATTATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 4260
Db      4261 aattggatgaataatgtagaataatgtagaataatgtagaataatgtagaataatgtaga 4320
Qy      4261 AATTGGATGATCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Db      4321 tcaacgatctctctctcctgagatagatgtagaataatgtagaataatgtagaataatgtaga 4380
Qy      4321 TCACGGATCTCTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4380
Db      4381 ctgagcaactcaatcatatcatatgtagaataatgtagaataatgtagaataatgtagaataatgtaga 4440
Qy      4381 CTGAGCAACTCAATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATAT 4440
Db      4441 taagcgtgttttcgagcaggaataacatgtagaataatgtagaataatgtagaataatgtaga 4500
Qy      4441 TAAGCGTGTGTTTCGAGCAGGAATAACATGATGATGATGATGATGATGATGATGATGATGAT 4500
Db      4501 gcccatcataaataatgtagaataatgtagaataatgtagaataatgtagaataatgtagaataatgtaga 4560
Qy      4501 GCCTCATCAATAATATGAAACATTTGACGCAATGCCACACGCAATGATACCAAGCCG 4560
Db      4561 acaagacaactcttacaactttgtagaataatgtagaataatgtagaataatgtagaataatgtaga 4600
Qy      4561 ACAAGACAATCTTACCAATTTGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 4600

RESULT 12
ID      054666 standard; DNA; 7653 BP.
AC      054666;
DT      23-JUN-1994 (first entry)
DE      ced-3 gene.
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain; ss.
OS      Caenorhabditis elegans.
FH      key location/Qualifiers
FT      repeat_unit 1356..1472

```

```

Db 3601 tcatgtggttttccatgaatgacccaaaatagcaattttatatatttcccrat 3660
QY 3601 tcatgtggttttccatgaatgacccaaaatagcaattttatatatttcccrat 3660
Db 3661 tcatgtgtgacgaagaaatagtaaaagcgatcatcttttcgacatttttaccatga 3720
QY 3661 tcatgtgtgacgaagaaatagtaaaagcgatcatcttttcgacatttttaccatga 3720
Db 3661 tcatgtgtgacgaagaaatagtaaaagcgatcatcttttcgacatttttaccatga 3720
QY 3661 tcatgtgtgacgaagaaatagtaaaagcgatcatcttttcgacatttttaccatga 3720
Db 3721 acgacagctcacttcaatcgtcgaagacgagagacgaggaataacacacatcttct 3780
QY 3721 acgacagctcacttcaatcgtcgaagacgagagacgaggaataacacacatcttct 3780
Db 3781 ggcgtctctgcttccgacatgtaaaatggatctggtcgtatgtaaaatgtaaaatg 3840
QY 3781 ggcgtctctgcttccgacatgtaaaatggatctggtcgtatgtaaaatgtaaaatg 3840
Db 3841 atgtaaaaatgcatgctgttttttaccatttctcgaacaaatagaggggaaatgt 3900
QY 3841 atgtaaaaatgcatgctgttttttaccatttctcgaacaaatagaggggaaatgt 3900
Db 3901 attaaatcatatttttgcatttttcaacatcagatgataaaccatatttttgcgt 3960
QY 3901 attaaatcatatttttgcatttttcaacatcagatgataaaccatatttttgcgt 3960
Db 3961 gggcaacttaaaagtagagaaatattagagcgaaacccaaatcttcaagatattcc 4020
QY 3961 gggcaacttaaaagtagagaaatattagagcgaaacccaaatcttcaagatattcc 4020
Db 4021 ttatgtataatattagatgtaataaagcatacttgaatgaagtcagcaaaatagt 4080
QY 4021 ttatgtataatattagatgtaataaagcatacttgaatgaagtcagcaaaatagt 4080
Db 4081 gggaaacacctgaaaaaaatcaaaaatctgcgaaaattgaaaaaatgcatataatca 4140
QY 4081 gggaaacacctgaaaaaaatcaaaaatctgcgaaaattgaaaaaatgcatataatca 4140
Db 4141 tttttcatttttccatcagatgagatgtagaaatgaaggaatcaaatcttcta 4200
QY 4141 tttttcatttttccatcagatgagatgtagaaatgaaggaatcaaatcttcta 4200
Db 4201 gaggatataatgtaaacatctgcgaatataaagtgcgaacacgtaaaaaaagaga 4260
QY 4201 gaggatataatgtaaacatctgcgaatataaagtgcgaacacgtaaaaaaagaga 4260
Db 4261 aatttgggtatcaaaatcgatccaaacacacatctcagcatccgcacacttcat 4320
QY 4261 aatttgggtatcaaaatcgatccaaacacacatctcagcatccgcacacttcat 4320
Db 4321 tccacggatgctcttctcgcgatacagttcagatcgtcattcagcaagctt 4380
QY 4321 tccacggatgctcttctcgcgatacagttcagatcgtcattcagcaagctt 4380
Db 4381 ctggacaaactcaatcatatccatgaaagagatatagaactttgcatgacacacca 4440
QY 4381 ctggacaaactcaatcatatccatgaaagagatatagaactttgcatgacacacca 4440
Db 4441 taagccggttttttcgacgagaaacatgtacagaaacttccgagtcctgtaagt 4500
QY 4441 taagccggttttttcgacgagaaacatgtacagaaacttccgagtcctgtaagt 4500
Db 4501 ggcctacataaataatgacacattgagacagatgacagagatgacagagcg 4560
QY 4501 ggcctacataaataatgacacattgagacagatgacagagatgacagagcg 4560
Db 4561 acaagacaatcttaccatattgttcagatgagctga 4600
QY 4561 acaagacaatcttaccatattgttcagatgagctga 4600

```

RESULT 11
ID 064740 standard; DNA: 7653 BP.
AC 064740;

```

DT 23-JUN-1994 (first entry)
DE ced-3 (C6322F) gene.
KW C.elegans; ced-4; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT /tag= a
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT FT repeat_unit 1490..1614
FT FT /tag= b
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT repeat_unit 2232..2366
FT FT /tag= c
FT FT /number= Exon_1
FT FT /tag= d
FT FT /number= Intron_1
FT FT /tag= e
FT FT /number= Exon_2
FT FT /tag= f
FT FT /number= Intron_2
FT FT /tag= g
FT FT /number= Exon_3
FT FT /tag= h
FT FT /number= Intron_3
FT FT /tag= i
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT repeat_unit 3329..3396
FT FT /tag= j
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT repeat_unit 3487..3759
FT FT /tag= k
FT FT /rpl_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT repeat_unit 4303..4634
FT FT /tag= m
FT FT /number= Exon_4
FT FT /tag= n
FT FT /number= Intron_4
FT FT /tag= o
FT FT /rpl_type= INVERTED
FT FT /note= "Repeat 3"
FT FT repeat_unit 5221..5330
FT FT /tag= p
FT FT /rpl_type= INVERTED
FT FT /note= "Repeat 3"
FT FT exon 5347..5760
FT FT /tag= q
FT FT /number= Exon_5
FT FT /tag= r
FT FT /number= Intron_5
FT FT /tag= s
FT FT /number= Exon_6

```

```

DB 4561 acaagacatctaccattgttcacatgcacatgcta 4600
OY 4561 ACAAGGACATCTTACCATTTGTCAGATGCATGGCTA 4600

RESULT 10
ID 064743 standard; DNA: 7653 BP.
AC 064743: 1994 (first entry)
DE ced-3 (C64857) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FH Repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /tag= g
FT /number= Exon_3
FT intron 3108..4302
FT /tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /tag= m
FT /number= Exon_4
FT intron 4635..5546
FT /tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"

```

```

FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT repeat_region 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT mutation 6485
FT /tag= w
FT /note= "C>T, from allele n2430"
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT exon 7012..7075
FT /number= Exon_8
PN W09325685-A.
PD 23-DEC-1993.
PE 14-JUN-1993. U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-897638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shiham S, Yuan J.
DR WPI, 94-007542/01.
DR P-PSDB; R33286.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in 064735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

FT      /number= Intron_4
FT      4688..4719
FT      /tag= O
FT      /rpl_type= INVERTED
FT      /note= "Repeat 3"
FT      5221..5330
FT      /tag= P
FT      /rpl_type= INVERTED
FT      /note= Repeat 3"
FT      5347..5760
FT      /tag= q
FT      /number= Exon_5
FT      5761..5814
FT      /tag= r
FT      /number= Intron_5
FT      5815..5942
FT      /tag= s
FT      /number= Exon_6
FT      5943..6297
FT      /tag= t
FT      /number= Intron_6
FT      6062..6138
FT      /tag= u
FT      /rpl_type= INVERTED
FT      6298..6537
FT      /tag= v
FT      /number= Exon_7
FT      6434
FT      /tag= w
FT      /note= "C>T, from allele n1129 and n1164"
FT      6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      6567..6625
FT      /tag= y
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      /tag= z
FT      /rpl_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7012..7075
FT      /tag= aa
FT      /number= Exon_8
FT      7076..7101
FT      /tag= ab
FT      /number= Exon_9
FT      7102..7127
FT      /tag= ac
FT      /number= Exon_10
FT      7128..7153
FT      /tag= ad
FT      /number= Exon_11
FT      7154..7179
FT      /tag= ae
FT      /number= Exon_12
FT      7180..7205
FT      /tag= af
FT      /number= Exon_13
FT      7206..7231
FT      /tag= ag
FT      /number= Exon_14
FT      7232..7257
FT      /tag= ah
FT      /number= Exon_15
FT      7258..7283
FT      /tag= ai
FT      /number= Exon_16
FT      7284..7309
FT      /tag= aj
FT      /number= Exon_17
FT      7310..7335
FT      /tag= ak
FT      /number= Exon_18
FT      7336..7361
FT      /tag= al
FT      /number= Exon_19
FT      7362..7387
FT      /tag= am
FT      /number= Exon_20
FT      7388..7413
FT      /tag= an
FT      /number= Exon_21
FT      7414..7439
FT      /tag= ao
FT      /number= Exon_22
FT      7440..7465
FT      /tag= ap
FT      /number= Exon_23
FT      7466..7491
FT      /tag= aq
FT      /number= Exon_24
FT      7492..7517
FT      /tag= ar
FT      /number= Exon_25
FT      7518..7543
FT      /tag= as
FT      /number= Exon_26
FT      7544..7569
FT      /tag= at
FT      /number= Exon_27
FT      7570..7595
FT      /tag= au
FT      /number= Exon_28
FT      7596..7621
FT      /tag= av
FT      /number= Exon_29
FT      7622..7647
FT      /tag= aw
FT      /number= Exon_30
FT      7648..7673
FT      /tag= ax
FT      /number= Exon_31
FT      7674..7699
FT      /tag= ay
FT      /number= Exon_32
FT      7700..7725
FT      /tag= az
FT      /number= Exon_33
FT      7726..7751
FT      /tag= ba
FT      /number= Exon_34
FT      7752..7777
FT      /tag= bb
FT      /number= Exon_35
FT      7778..7803
FT      /tag= bc
FT      /number= Exon_36
FT      7804..7829
FT      /tag= bd
FT      /number= Exon_37
FT      7830..7855
FT      /tag= be
FT      /number= Exon_38
FT      7856..7881
FT      /tag= bf
FT      /number= Exon_39
FT      7882..7907
FT      /tag= bg
FT      /number= Exon_40
FT      7908..7933
FT      /tag= bh
FT      /number= Exon_41
FT      7934..7959
FT      /tag= bi
FT      /number= Exon_42
FT      7960..7985
FT      /tag= bj
FT      /number= Exon_43
FT      7986..8011
FT      /tag= bk
FT      /number= Exon_44
FT      8012..8037
FT      /tag= bl
FT      /number= Exon_45
FT      8038..8063
FT      /tag= bm
FT      /number= Exon_46
FT      8064..8089
FT      /tag= bn
FT      /number= Exon_47
FT      8090..8115
FT      /tag= bo
FT      /number= Exon_48
FT      8116..8141
FT      /tag= bp
FT      /number= Exon_49
FT      8142..8167
FT      /tag= bq
FT      /number= Exon_50
FT      8168..8193
FT      /tag= br
FT      /number= Exon_51
FT      8194..8219
FT      /tag= bs
FT      /number= Exon_52
FT      8220..8245
FT      /tag= bt
FT      /number= Exon_53
FT      8246..8271
FT      /tag= bu
FT      /number= Exon_54
FT      8272..8297
FT      /tag= bv
FT      /number= Exon_55
FT      8298..8323
FT      /tag= bw
FT      /number= Exon_56
FT      8324..8349
FT      /tag= bx
FT      /number= Exon_57
FT      8350..8375
FT      /tag= by
FT      /number= Exon_58
FT      8376..8401
FT      /tag= bz
FT      /number= Exon_59
FT      8402..8427
FT      /tag= ca
FT      /number= Exon_60
FT      8428..8453
FT      /tag= cb
FT      /number= Exon_61
FT      8454..8479
FT      /tag= cc
FT      /number= Exon_62
FT      8480..8505
FT      /tag= cd
FT      /number= Exon_63
FT      8506..8531
FT      /tag= ce
FT      /number= Exon_64
FT      8532..8557
FT      /tag= cf
FT      /number= Exon_65
FT      8558..8583
FT      /tag= cg
FT      /number= Exon_66
FT      8584..8609
FT      /tag= ch
FT      /number= Exon_67
FT      8610..8635
FT      /tag= ci
FT      /number= Exon_68
FT      8636..8661
FT      /tag= cj
FT      /number= Exon_69
FT      8662..8687
FT      /tag= ck
FT      /number= Exon_70
FT      8688..8713
FT      /tag= cl
FT      /number= Exon_71
FT      8714..8739
FT      /tag= cm
FT      /number= Exon_72
FT      8740..8765
FT      /tag= cn
FT      /number= Exon_73
FT      8766..8791
FT      /tag= co
FT      /number= Exon_74
FT      8792..8817
FT      /tag= cp
FT      /number= Exon_75
FT      8818..8843
FT      /tag= cq
FT      /number= Exon_76
FT      8844..8869
FT      /tag= cr
FT      /number= Exon_77
FT      8870..8895
FT      /tag= cs
FT      /number= Exon_78
FT      8896..8921
FT      /tag= ct
FT      /number= Exon_79
FT      8922..8947
FT      /tag= cu
FT      /number= Exon_80
FT      8948..8973
FT      /tag= cv
FT      /number= Exon_81
FT      8974..9000
FT      /tag= cw
FT      /number= Exon_82
FT      9001..9026
FT      /tag= cx
FT      /number= Exon_83
FT      9027..9052
FT      /tag= cy
FT      /number= Exon_84
FT      9053..9078
FT      /tag= cz
FT      /number= Exon_85
FT      9079..9104
FT      /tag= da
FT      /number= Exon_86
FT      9105..9130
FT      /tag= db
FT      /number= Exon_87
FT      9131..9156
FT      /tag= dc
FT      /number= Exon_88
FT      9157..9182
FT      /tag= dd
FT      /number= Exon_89
FT      9183..9208
FT      /tag= de
FT      /number= Exon_90
FT      9209..9234
FT      /tag= df
FT      /number= Exon_91
FT      9235..9260
FT      /tag= dg
FT      /number= Exon_92
FT      9261..9286
FT      /tag= dh
FT      /number= Exon_93
FT      9287..9312
FT      /tag= di
FT      /number= Exon_94
FT      9313..9338
FT      /tag= dj
FT      /number= Exon_95
FT      9339..9364
FT      /tag= dk
FT      /number= Exon_96
FT      9365..9390
FT      /tag= dl
FT      /number= Exon_97
FT      9391..9416
FT      /tag= dm
FT      /number= Exon_98
FT      9417..9442
FT      /tag= dn
FT      /number= Exon_99
FT      9443..9468
FT      /tag= do
FT      /number= Exon_100
FT      9469..9494
FT      /tag= dp
FT      /number= Exon_101
FT      9495..9520
FT      /tag= dq
FT      /number= Exon_102
FT      9521..9546
FT      /tag= dr
FT      /number= Exon_103
FT      9547..9572
FT      /tag= ds
FT      /number= Exon_104
FT      9573..9598
FT      /tag= dt
FT      /number= Exon_105
FT      9599..9624
FT      /tag= du
FT      /number= Exon_106
FT      9625..9650
FT      /tag= dv
FT      /number= Exon_107
FT      9651..9676
FT      /tag= dw
FT      /number= Exon_108
FT      9677..9702
FT      /tag= dx
FT      /number= Exon_109
FT      9703..9728
FT      /tag= dy
FT      /number= Exon_110
FT      9729..9754
FT      /tag= dz
FT      /number= Exon_111
FT      9755..9780
FT      /tag= ea
FT      /number= Exon_112
FT      9781..9806
FT      /tag= eb
FT      /number= Exon_113
FT      9807..9832
FT      /tag= ec
FT      /number= Exon_114
FT      9833..9858
FT      /tag= ed
FT      /number= Exon_115
FT      9859..9884
FT      /tag= ee
FT      /number= Exon_116
FT      9885..9910
FT      /tag= ef
FT      /number= Exon_117
FT      9911..9936
FT      /tag= eg
FT      /number= Exon_118
FT      9937..9962
FT      /tag= eh
FT      /number= Exon_119
FT      9963..9988
FT      /tag= ei
FT      /number= Exon_120
FT      9989..10014
FT      /tag= ej
FT      /number= Exon_121
FT      10015..10040
FT      /tag= ek
FT      /number= Exon_122
FT      10041..10066
FT      /tag= el
FT      /number= Exon_123
FT      10067..10092
FT      /tag= em
FT      /number= Exon_124
FT      10093..10118
FT      /tag= en
FT      /number= Exon_125
FT      10119..10144
FT      /tag= eo
FT      /number= Exon_126
FT      10145..10170
FT      /tag= ep
FT      /number= Exon_127
FT      10171..10196
FT      /tag= eq
FT      /number= Exon_128
FT      10197..10222
FT      /tag= er
FT      /number= Exon_129
FT      10223..10248
FT      /tag= es
FT      /number= Exon_130
FT      10249..10274
FT      /tag= et
FT      /number= Exon_131
FT      10275..10300
FT      /tag= eu
FT      /number= Exon_132
FT      10301..10326
FT      /tag= ev
FT      /number= Exon_133
FT      10327..10352
FT      /tag= ew
FT      /number= Exon_134
FT      10353..10378
FT      /tag= ex
FT      /number= Exon_135
FT      10379..10404
FT      /tag= ey
FT      /number= Exon_136
FT      10405..10430
FT      /tag= ez
FT      /number= Exon_137
FT      10431..10456
FT      /tag= fa
FT      /number= Exon_138
FT      10457..10482
FT      /tag= fb
FT      /number= Exon_139
FT      10483..10508
FT      /tag= fc
FT      /number= Exon_140
FT      10509..10534
FT      /tag= fd
FT      /number= Exon_141
FT      10535..10560
FT      /tag= fe
FT      /number= Exon_142
FT      10561..10586
FT      /tag= ff
FT      /number= Exon_143
FT      10587..10612
FT      /tag= fg
FT      /number= Exon_144
FT      10613..10638
FT      /tag= fh
FT      /number= Exon_145
FT      10639..10664
FT      /tag= fi
FT      /number= Exon_146
FT      10665..10690
FT      /tag= fj
FT      /number= Exon_147
FT      10691..10716
FT      /tag= fk
FT      /number= Exon_148
FT      10717..10742
FT      /tag= fl
FT      /number= Exon_149
FT      10743..10768
FT      /tag= fm
FT      /number= Exon_150
FT      10769..10794
FT      /tag= fn
FT      /number= Exon_151
FT      10795..10820
FT      /tag= fo
FT      /number= Exon_152
FT      10821..10846
FT      /tag= fp
FT      /number= Exon_153
FT      10847..10872
FT      /tag= fq
FT      /number= Exon_154
FT      10873..10898
FT      /tag= fr
FT      /number= Exon_155
FT      10899..10924
FT      /tag= fs
FT      /number= Exon_156
FT      10925..10950
FT      /tag= ft
FT      /number= Exon_157
FT      10951..10976
FT      /tag= fu
FT      /number= Exon_158
FT      10977..11002
FT      /tag= fv
FT      /number= Exon_159
FT      11003..11028
FT      /tag= fw
FT      /number= Exon_160
FT      11029..11054
FT      /tag= fx
FT      /number= Exon_161
FT      11055..11080
FT      /tag= fy
FT      /number= Exon_162
FT      11081..11106
FT      /tag= fz
FT      /number= Exon_163
FT      11107..11132
FT      /tag= ga
FT      /number= Exon_164
FT      11133..11158
FT      /tag= gb
FT      /number= Exon_165
FT      11159..11184
FT      /tag= gc
FT      /number= Exon_166
FT      11185..11210
FT      /tag= gd
FT      /number= Exon_167
FT      11211..11236
FT      /tag= ge
FT      /number= Exon_168
FT      11237..11262
FT      /tag= gf
FT      /number= Exon_169
FT      11263..11288
FT      /tag= gh
FT      /number= Exon_170
FT      11289..11314
FT      /tag= gi
FT      /number= Exon_171
FT      11315..11340
FT      /tag= gj
FT      /number= Exon_172
FT      11341..11366
FT      /tag= gk
FT      /number= Exon_173
FT      11367..11392
FT      /tag= gl
FT      /number= Exon_174
FT      11393..11418
FT      /tag= gm
FT      /number= Exon_175
FT      11419..11444
FT      /tag= gn
FT      /number= Exon_176
FT      11445..11470
FT      /tag= go
FT      /number= Exon_177
FT      11471..11496
FT      /tag= gp
FT      /number= Exon_178
FT      11497..11522
FT      /tag= gq
FT      /number= Exon_179
FT      11523..11548
FT      /tag= gr
FT      /number= Exon_180
FT      11549..11574
FT      /tag= gs
FT      /number= Exon_181
FT      11575..11600
FT      /tag= gt
FT      /number= Exon_182
FT      11601..11626
FT      /tag= gu
FT      /number= Exon_183
FT      11627..11652
FT      /tag= gv
FT      /number= Exon_184
FT      11653..11678
FT      /tag= gw
FT      /number= Exon_185
FT      11679..11704
FT      /tag= gx
FT      /number= Exon_186
FT      11705..11730
FT      /tag= gy
FT      /number= Exon_187
FT      11731..11756
FT      /tag= gz
FT      /number= Exon_188
FT      11757..11782
FT      /tag= ha
FT      /number= Exon_189
FT      11783..11808
FT      /tag= hb
FT      /number= Exon_190
FT      11809..11834
FT      /tag= hc
FT      /number= Exon_191
FT      11835..11860
FT      /tag= hd
FT      /number= Exon_192
FT      11861..11886
FT      /tag= he
FT      /number= Exon_193
FT      11887..11912
FT      /tag= hf
FT      /number= Exon_194
FT      11913..11938
FT      /tag= hg
FT      /number= Exon_195
FT      11939..11964
FT      /tag= hh
FT      /number= Exon_196
FT      11965..11990
FT      /tag= hi
FT      /number= Exon_197
FT      11991..12016
FT      /tag= hj
FT      /number= Exon_198
FT      12017..12042
FT      /tag= hk
FT      /number= Exon_199
FT      12043..12068
FT      /tag= hl
FT      /number= Exon_200
FT      12069..12094
FT      /tag= hm
FT      /number= Exon_201
FT      12095..12120
FT      /tag= hn
FT      /number= Exon_202
FT      12121..12146
FT      /tag= ho
FT      /number= Exon_203
FT      12147..12172
FT      /tag= hp
FT      /number= Exon_204
FT      12173..12198
FT      /tag= hq
FT      /number= Exon_205
FT      12199..12224
FT      /tag= hr
FT      /number= Exon_206
FT      12225..12250
FT      /tag= hs
FT      /number= Exon_207
FT      12251..12276
FT      /tag= ht
FT      /number= Exon_208
FT      12277..12302
FT      /tag= hu
FT      /number= Exon_209
FT      12303..12328
FT      /tag= hv
FT      /number= Exon_210
FT      12329..12354
FT      /tag= hw
FT      /number= Exon_211
FT      12355..12380
FT      /tag= hx
FT      /number= Exon_212
FT      12381..12406
FT      /tag= hy
FT      /number= Exon_213
FT      12407..12432
FT      /tag= hz
FT      /number= Exon_214
FT      12433..12458
FT      /tag= ia
FT      /number= Exon_215
FT      12459..12484
FT      /tag= ib
FT      /number= Exon_216
FT      12485..12510
FT      /tag= ic
FT      /number= Exon_217
FT      12511..12536
FT      /tag= id
FT      /number= Exon_218
FT      12537..12562
FT      /tag= ie
FT      /number= Exon_219
FT      12563..12588
FT      /tag= if
FT      /number= Exon_220
FT      12589..12614
FT      /tag= ig
FT      /number= Exon_221
FT      12615..12640
FT      /tag= ih
FT      /number= Exon_222
FT      12641..12666
FT      /tag= ii
FT      /number= Exon_223
FT      12667..12692
FT      /tag= ij
FT      /number= Exon_224
FT      12693..12718
FT      /tag= ik
FT      /number= Exon_225
FT      12719..12744
FT      /tag= il
FT      /number= Exon_226
FT      12745..12770
FT      /tag= im
FT      /number= Exon_227
FT      12771..12796
FT      /tag= in
FT      /number= Exon_228
FT      12797..12822
FT      /tag= io
FT      /number= Exon_229
FT      12823..12848
FT      /tag= ip
FT      /number= Exon_230
FT      12849..12874
FT      /tag= iq
FT      /number= Exon_231
FT      12875..12900
FT      /tag= ir
FT      /number= Exon_232
FT      12901..12926
FT      /tag= is
FT      /number= Exon_233
FT      12927..12952
FT      /tag= it
FT      /number= Exon_234
FT      12953..12978
FT      /tag= iv
FT      /number= Exon_235
FT      12979..13004
FT      /tag= iw
FT      /number= Exon_236
FT      13005..13030
FT      /tag= ix
FT      /number= Exon_237
FT      13031..13056
FT      /tag= iy
FT      /number= Exon_238
FT      13057..13082
FT      /tag= iz
FT      /number= Exon_239
FT      13083..13108
FT      /tag= ja
FT      /number= Exon_240
FT      13109..13134
FT      /tag= jb
FT      /number= Exon_241
FT      13135..13160
FT      /tag= jc
FT      /number= Exon_242
FT      13161..13186
FT      /tag= jd
FT      /number= Exon_243
FT      13187..13212
FT      /tag= je
FT      /number= Exon_244
FT      13213..13238
FT      /tag= jf
FT      /number= Exon_245
FT      13239..13264
FT      /tag= jg
FT      /number= Exon_246
FT      13265..13290
FT      /tag= jh
FT      /number= Exon_247
FT      13291..13316
FT      /tag= ji
FT      /number= Exon_248
FT      13317..13342
FT      /tag= jj
FT      /number= Exon_249
FT      13343..13368
FT      /tag= jk
FT      /number= Exon_250
FT      13369..13394
FT      /tag= jl
FT      /number= Exon_251
FT      13395..13420
FT      /tag= jm
FT      /number= Exon_252
FT      13421..13446
FT      /tag= jn
FT      /number= Exon_253
FT      13447..13472
FT      /tag= jo
FT      /number= Exon_254
FT      13473..13498
FT      /tag= jp
FT      /number= Exon_255
FT      13499..13524
FT      /tag= jq
FT      /number= Exon_256
FT      13525..13550
FT      /tag= jr
FT      /number= Exon_257
FT      13551..13576
FT      /tag= js
FT      /number= Exon_258
FT      13577..13602
FT      /tag= jt
FT      /number= Exon_259
FT      13603..13628
FT      /tag= ju
FT      /number= Exon_260
FT      13629..13654
FT      /tag= jv
FT      /number= Exon_261
FT      13655..13680
FT      /tag= jw
FT      /number= Exon_262
FT      13681..13706
FT      /tag= jx
FT      /number= Exon_263
FT      13707..13732
FT      /tag= jy
FT      /number= Exon_264
FT      13733..13758
FT      /tag= jz
FT      /number= Exon_265
FT      13759..13784
FT      /tag= ka
FT      /number= Exon_266
FT      13785..13810
FT      /tag= kb
FT      /number= Exon_267
FT      13811..13836
FT      /tag= kc
FT      /number= Exon_268
FT      13837..13862
FT      /tag= kd
FT      /number= Exon_269
FT      13863..13888
FT      /tag= ke
FT      /number= Exon_270
FT      13889..13914
FT      /tag= kf
FT      /number= Exon_271
FT      13915..13940
FT      /tag= kg
FT      /number= Exon_272
FT      13941..13966
FT      /tag= kh
FT      /number= Exon_273
FT      13967..13992
FT      /tag= ki
FT      /number= Exon_274
FT      13993..14018
FT      /tag= kj
FT      /number= Exon_275
FT      14019..14044
FT      /tag= kk
FT      /number= Exon_276
FT      14045..14070
FT      /tag= kl
FT      /number= Exon_277
FT      14071..14096
FT      /tag= km
FT      /number= Exon_278
FT      14097..14122
FT      /tag= kn
FT      /number= Exon_279
FT      14123..14148
FT      /tag= ko
FT      /number= Exon_280
FT      14149..14174
FT      /tag= kp
FT      /number= Exon_281
FT      14175..14200
FT      /tag= kq
FT      /number= Exon_282
FT      14201..14226
FT      /tag= kr
FT      /number= Exon_283
FT      14227..14252
FT      /tag= ks
FT      /number= Exon_284
FT      14253..14278
FT      /tag= kt
FT      /number= Exon_285
FT      14279..14304
FT      /tag= ku
FT      /number= Exon_286
FT      14305..14330
FT      /tag= kv
FT      /number= Exon_287
FT      14331..14356
FT      /tag= kw
FT      /number= Exon_288
FT      14357..14382
FT      /tag= kx
FT      /number= Exon_289
FT      14383..14408
FT      /tag= ky
FT      /number= Exon_290
FT      14409..14434
FT      /tag= kz
FT      /number= Exon_291
FT      14435..14460
FT      /tag= la
FT      /number= Exon_292
FT      14461..14486
FT      /tag= lb
FT      /number= Exon_293
FT      14487..14512
FT      /tag= lc
FT      /number= Exon_294
FT      14513..14538
FT      /tag= ld
FT      /number= Exon_295
FT      14539..14564
FT      /tag= le
FT      /number= Exon_296
FT      14565..14590
FT      /tag= lf
FT      /number= Exon_297
FT      14591..14616
FT      /tag= lg
FT      /number= Exon_298
FT      14617..14642
FT      /tag= lh
FT      /number= Exon_299
FT      14643..14668
FT      /tag= li
FT      /number= Exon_300
FT      14669..14694
FT      /tag= lj
FT      /number= Exon_301
FT      14695..14720
FT      /tag= lk
FT      /number= Exon_302
FT      14721..14746
FT      /tag= ll
FT      /number= Exon_303
FT      14747..14772
FT      /tag= lm
FT      /number= Exon_304
FT      14773..14798
FT      /tag= ln
FT      /number= Exon_305
FT      14799..14824
FT      /tag= lo
FT      /number= Exon_306
FT      14825..14850
FT      /tag= lp
FT      /number= Exon_307
FT      14851..14876
FT      /tag= lp
FT      /number= Exon_308
FT      14877..14902
FT      /tag= lr
FT      /number= Exon_309
FT      14903..14928
FT      /tag= ls
FT      /number= Exon_310
FT      14929..14954
FT      /tag= lt
FT      /number= Exon_311
FT      14955..14980
FT      /tag= lu
FT      /number= Exon_312
FT      14981..15006
FT      /tag= lv
FT      /number= Exon_313
FT      15007..15032
FT      /tag= lw
FT      /number= Exon_314
FT      15033..15058
FT      /tag= lx
FT      /number= Exon_315
FT      15059..15084
FT      /tag= ly
FT      /number= Exon_316
FT      15085..15110
FT      /tag= lz
FT      /number= Exon_317
FT      15111..15136
FT      /tag= ma
FT      /number= Exon_318
FT      15137..15162
FT      /tag= mb
FT      /number= Exon_319
FT      15163..15188
FT      /tag= mc
FT      /number= Exon_320
FT      15189..15214
FT      /tag= md
FT      /number= Exon_321
FT      15215..15240
FT      /tag= me
FT      /number= Exon_322
FT      15241..15266
FT      /tag= mf
FT      /number= Exon_323
FT      15267..15292
FT      /tag= mg
FT      /number= Exon_324
FT      15293..15318
FT      /tag= mh
FT      /number= Exon_325
FT      15319..15344
FT      /tag= mi
FT      /number= Exon_326
FT      15345..15370
FT      /tag= mj
FT      /number= Exon_327
FT      15371..15396
FT      /tag= mk
FT      /number= Exon_328
FT      15397..15422
FT      /tag= ml
FT      /number= Exon_329
FT      15423..15448
FT      /tag= mn
FT      /number= Exon_330
FT      15449..15474
FT      /tag= mo
FT      /number= Exon_331
FT      15475..15500
FT      /tag= mp
FT      /number= Exon_332
FT      15501..15526
FT      /tag= mq
FT      /number= Exon_333
FT      15527..15552
FT      /tag= mr
FT      /number= Exon_334
FT      15553..15578
FT      /tag= ms
FT      /number= Exon_335
FT      15579..15604
FT      /tag= mt
FT      /number= Exon_336
FT      15605..15630
FT      /tag= mu
FT      /number= Exon_337
FT      15631..15656
FT      /tag= mv
FT      /number= Exon_338
FT      15657..15682
FT      /tag= mw
FT      /number= Exon_339
FT      15683..15708
FT      /tag= mx
FT      /number= Exon_340
FT      15709..15734
FT      /tag= my
FT      /number= Exon_341
FT      15735..15760
FT      /tag= mz
FT      /number= Exon_342
FT      15761..15786
FT      /tag= na
FT      /number= Exon_343
FT      15787..15812
FT      /tag= nb
FT      /number= Exon_344
FT      15813..15838
FT      /tag= nc
FT      /number= Exon_345
FT      15839..15864
FT      /tag= nd
FT      /number= Exon_346
FT      15865..15890
FT      /tag= ne
FT      /number= Exon_347
FT      15891..15916
FT      /tag= nf
FT      /number= Exon_348
FT      15917..15942
FT      /tag= ng
FT      /number= Exon_349
FT      15943..15968
FT      /tag= nh
FT      /number= Exon_350
FT      15969..15994
FT      /tag= ni
FT      /number= Exon_351
FT      15995..16020
FT      /tag= nj
FT      /number= Exon_352
FT      16021..16046
FT      /tag= nk
FT      /number= Exon_353
FT      16047..16072
FT      /tag= nl
FT      /number= Exon_354
FT      16073..16098
FT      /tag= nm
FT      /number= Exon_355
FT      16099..16124
FT      /tag= nn
FT      /number= Exon_356
FT      16125..16150
FT      /tag= no
FT      /number= Exon_357
FT      16151..16176
FT      /tag= np
FT      /number= Exon_358
FT      16177..16202
FT      /tag= nq
FT      /number= Exon_359
FT      16203..16228
FT      /tag= nr
FT      /number= Exon_360
FT      16229..16254
FT      /tag= ns
FT      /number= Exon_361
FT      16255..16280
FT      /tag= nt
FT      /number= Ex
```

CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations; two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;
 Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3601 tcatgtggttgcataatgaatgacccaaatagcaatttttatataattccctat 3660
 |||||||
 Qy 3601 TCATGCGTTGTTCATATGATGACGAAATAGCAATTTTATATATTTCCCTAT 3660
 |||||||
 Db 3661 tcatgttgagcaaaaagaagaagcgatcatcttttcgacatcttttaccatga 3720
 |||||||
 Qy 3661 TCATGTTGTGCAAAAAATGTAAGAACCGATCATTTTTCGACATTTTTCATGCA 3720
 |||||||
 Db 3721 acgacagccatctcacaatgcagagacgcgcgagagaatacacacatcttct 3780
 |||||||
 Qy 3721 AGCAGACGCTCATCTGCATGCTGAAGAGACGAGACGCGGAGAAATACACACATCTTCT 3780
 |||||||
 Db 3781 gcgtctctgcgttcagcaatggaatggatcgcgtcgtatgtataaaatgctgata 3840
 |||||||
 Qy 3781 CGGTCTCTGCTTGCATGTAATGGAATGGATCGGTGATGTAATAAATGTGAAAT 3840
 |||||||
 Db 3841 atgtataaaatgcagcgttttttaccacttttcgcacaaatagaggggaaatcgt 3900
 |||||||
 Qy 3841 ATGTATAAATGCATGCGGTTTTCACCTTTTTCGACAAATGATAGGGGAAATGT 3900
 |||||||
 Db 3901 attaaatcatcttttgcatttttcaacatcacatgatacccatatttttcgt 3960
 |||||||
 Qy 3901 ATTAAATCATTTTGTATTTTTCACATCATCATGATTAACCCCATATTTTTCGT 3960
 |||||||
 Db 3961 gagcaactaaagtagaataatagagcgaaacccaatttccttaagatatacc 4020
 |||||||
 Qy 3961 GAGCAACTTAAAGTAGAATATAGAGCGAAACCAAAATTTCTCAAGATATAC 4020
 |||||||
 Db 4021 ttattgataattatagatggttaataagcatatcttgatgaagaatcagcaaaatagt 4080
 |||||||
 Qy 4021 TTATTGATTAATATAGATGTTAATAGCATATCTTGAATGAAGTACGCAAAATATGT 4080
 |||||||
 Db 4081 ggcgaacacctgaaaaaaatcgaatctgcgaatattgaaaaaaatgatatataa 4140
 |||||||
 Qy 4081 GCGAAACACCTGAAAAAATCAAAATTCGCGAAATTTGAAAAAATGATTAATAAATA 4140
 |||||||
 Db 4141 ttttgcatttttctcatcatcatgaaatgtagaaaattaaagggaaatcaaatctta 4200
 |||||||
 Qy 4141 TTTTTCATTTTCTACATCATCATGTAATGTAGAAATTTAAAGGGGAAATCAAAATTTCTA 4200
 |||||||
 Db 4201 gaggataatggaatgaacatctgcgaatataaaatgtagcaaaatgtaaaagaaga 4260
 |||||||
 Qy 4201 GAGGATTAATTAATGAAGAACATTTGCGAAATTAATTTGCGAAACGTAAATAAGAGA 4260
 |||||||
 Db 4261 aatttgatcaaaaatgcataacccaacacacattcagatccgcaactctcat 4320
 |||||||
 Qy 4261 AATTGGGATCAAAATCGATCTTAACCAACACATTTTCAGATTCGCGCACTCTTCA 4320
 |||||||
 Db 4321 tcacggatgctctctcgcagatcacagttcaagtcgtatcgtcattcagcaagctt 4380
 |||||||
 Qy 4321 TCACGGATGCTCTCTCGATACAGTTCAGTCAAGTCAATGCTCATTCAGCAAGCTT 4380
 |||||||
 Db 4381 ctggaccaactcaatcatattccatgagagagatatgaacttgcgtgacacaacca 4440
 |||||||

Qy 4381 CTGACCACTCAATATTCATTCATGAGAGATATGAACTTTGTGATGCAACCAACA 4440
 |||||||
 Db 4441 taagcggttttcgacgagaaaacatctacgaacttcgcagctcgtggaatg 4500
 |||||||
 Qy 4441 TAAGCGTGTTCGACGAGAAACCATGTACAGAAATCTTCGATCTCTCGTGAATGT 4500
 |||||||
 Db 4501 gcttcataaataatgacacttgagcagatgacacacggaatgttacaaggcgc 4560
 |||||||
 Qy 4501 GCTTCATCAATAATATGAACACTTTGACGAGATGCCAACGGAATGTACCAAGCCG 4560
 |||||||
 Db 4561 acaagacaatcttacaattgttcagatgcattgagcga 4600
 |||||||
 Qy 4561 ACAAGCAATCTTACCAATTTGTGATGATGATGATGATGATGATGATGATGATGAT 4600
 |||||||

RESULT 9
 ID 064742 standard; DNA; 7653 BP.
 AC 064742;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (C6434T) gene.
 KW C. elegans; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH key location/Qualifiers
 FT repeat_unit 1356..1472
 FT /*tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1490-1614"
 FT 1490..1614
 FT /*tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /*tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /*tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /*tag= e
 FT /number= Exon_2
 FT 2576..2853
 FT /*tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /*tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /*tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /*tag= i
 FT /number= Exon_4
 FT 3244..3396
 FT /*tag= j
 FT /number= Intron_4
 FT 3397..3546
 FT /*tag= k
 FT /number= Exon_5
 FT 3547..3759
 FT /*tag= l
 FT /number= Intron_5
 FT 3760..3914
 FT /*tag= m
 FT /number= Exon_6
 FT 3915..4069
 FT /*tag= n
 FT /number= Intron_6
 FT 4070..4224
 FT /*tag= o
 FT /number= Exon_7
 FT 4225..4379
 FT /*tag= p
 FT /number= Intron_7
 FT 4380..4534
 FT /*tag= q
 FT /number= Exon_8
 FT 4535..4689
 FT /*tag= r
 FT /number= Intron_8
 FT 4690..4844
 FT /*tag= s
 FT /number= Exon_9
 FT 4845..4999
 FT /*tag= t
 FT /number= Intron_9
 FT 5000..5154
 FT /*tag= u
 FT /number= Exon_10
 FT 5155..5309
 FT /*tag= v
 FT /number= Intron_10
 FT 5310..5464
 FT /*tag= w
 FT /number= Exon_11
 FT 5465..5619
 FT /*tag= x
 FT /number= Intron_11
 FT 5620..5774
 FT /*tag= y
 FT /number= Exon_12
 FT 5775..5929
 FT /*tag= z
 FT /number= Intron_12
 FT 5930..6084
 FT /*tag= AA
 FT /number= Exon_13
 FT 6085..6239
 FT /*tag= AB
 FT /number= Intron_13
 FT 6240..6394
 FT /*tag= AC
 FT /number= Exon_14
 FT 6395..6549
 FT /*tag= AD
 FT /number= Intron_14
 FT 6550..6704
 FT /*tag= AE
 FT /number= Exon_15
 FT 6705..6859
 FT /*tag= AF
 FT /number= Intron_15
 FT 6860..7014
 FT /*tag= AG
 FT /number= Exon_16
 FT 7015..7169
 FT /*tag= AH
 FT /number= Intron_16
 FT 7170..7324
 FT /*tag= AI
 FT /number= Exon_17
 FT 7325..7479
 FT /*tag= AJ
 FT /number= Intron_17
 FT 7480..7634
 FT /*tag= AK
 FT /number= Exon_18
 FT 7635..7789
 FT /*tag= AL
 FT /number= Intron_18
 FT 7790..7944
 FT /*tag= AM
 FT /number= Exon_19
 FT 7945..8099
 FT /*tag= AN
 FT /number= Intron_19
 FT 8100..8254
 FT /*tag= AO
 FT /number= Exon_20
 FT 8255..8409
 FT /*tag= AP
 FT /number= Intron_20
 FT 8410..8564
 FT /*tag= AQ
 FT /number= Exon_21
 FT 8565..8719
 FT /*tag= AR
 FT /number= Intron_21
 FT 8720..8874
 FT /*tag= AS
 FT /number= Exon_22
 FT 8875..9029
 FT /*tag= AT
 FT /number= Intron_22
 FT 9030..9184
 FT /*tag= AU
 FT /number= Exon_23
 FT 9185..9339
 FT /*tag= AV
 FT /number= Intron_23
 FT 9340..9494
 FT /*tag= AW
 FT /number= Exon_24
 FT 9495..9649
 FT /*tag= AX
 FT /number= Intron_24
 FT 9650..9804
 FT /*tag= AY
 FT /number= Exon_25
 FT 9805..9959
 FT /*tag= AZ
 FT /number= Intron_25
 FT 9960..10114
 FT /*tag= BA
 FT /number= Exon_26
 FT 10115..10269
 FT /*tag= BB
 FT /number= Intron_26
 FT 10270..10424
 FT /*tag= BC
 FT /number= Exon_27
 FT 10425..10579
 FT /*tag= BD
 FT /number= Intron_27
 FT 10580..10734
 FT /*tag= BE
 FT /number= Exon_28
 FT 10735..10889
 FT /*tag= BF
 FT /number= Intron_28
 FT 10890..11044
 FT /*tag= BG
 FT /number= Exon_29
 FT 11045..11199
 FT /*tag= BH
 FT /number= Intron_29
 FT 11200..11354
 FT /*tag= BI
 FT /number= Exon_30
 FT 11355..11509
 FT /*tag= BJ
 FT /number= Intron_30
 FT 11510..11664
 FT /*tag= BK
 FT /number= Exon_31
 FT 11665..11819
 FT /*tag= BL
 FT /number= Intron_31
 FT 11820..11974
 FT /*tag= BM
 FT /number= Exon_32
 FT 11975..12129
 FT /*tag= BN
 FT /number= Intron_32
 FT 12130..12284
 FT /*tag= BO
 FT /number= Exon_33
 FT 12285..12439
 FT /*tag= BP
 FT /number= Intron_33
 FT 12440..12594
 FT /*tag= BQ
 FT /number= Exon_34
 FT 12595..12749
 FT /*tag= BR
 FT /number= Intron_34
 FT 12750..12904
 FT /*tag= BS
 FT /number= Exon_35
 FT 12905..13059
 FT /*tag= BT
 FT /number= Intron_35
 FT 13060..13214
 FT /*tag= BU
 FT /number= Exon_36
 FT 13215..13369
 FT /*tag= BV
 FT /number= Intron_36
 FT 13370..13524
 FT /*tag= BW
 FT /number= Exon_37
 FT 13525..13679
 FT /*tag= BX
 FT /number= Intron_37
 FT 13680..13834
 FT /*tag= BY
 FT /number= Exon_38
 FT 13835..13989
 FT /*tag= BZ
 FT /number= Intron_38
 FT 13990..14144
 FT /*tag= CA
 FT /number= Exon_39
 FT 14145..14299
 FT /*tag= CB
 FT /number= Intron_39
 FT 14300..14454
 FT /*tag= CC
 FT /number= Exon_40
 FT 14455..14609
 FT /*tag= CD
 FT /number= Intron_40
 FT 14610..14764
 FT /*tag= CE
 FT /number= Exon_41
 FT 14765..14919
 FT /*tag= CF
 FT /number= Intron_41
 FT 14920..15074
 FT /*tag= CG
 FT /number= Exon_42
 FT 15075..15229
 FT /*tag= CH
 FT /number= Intron_42
 FT 15230..15384
 FT /*tag= CI
 FT /number= Exon_43
 FT 15385..15539
 FT /*tag= CJ
 FT /number= Intron_43
 FT 15540..15694
 FT /*tag= CK
 FT /number= Exon_44
 FT 15695..15849
 FT /*tag= CL
 FT /number= Intron_44
 FT 15850..16004
 FT /*tag= CM
 FT /number= Exon_45
 FT 16005..16159
 FT /*tag= CN
 FT /number= Intron_45
 FT 16160..16314
 FT /*tag= CO
 FT /number= Exon_46
 FT 16315..16469
 FT /*tag= CP
 FT /number= Intron_46
 FT 16470..16624
 FT /*tag= CQ
 FT /number= Exon_47
 FT 16625..16779
 FT /*tag= CR
 FT /number= Intron_47
 FT 16780..16934
 FT /*tag= CS
 FT /number= Exon_48
 FT 16935..17089
 FT /*tag= CT
 FT /number= Intron_48
 FT 17090..17244
 FT /*tag= CU
 FT /number= Exon_49
 FT 17245..17399
 FT /*tag= CV
 FT /number= Intron_49
 FT 17400..17554
 FT /*tag= CW
 FT /number= Exon_50
 FT 17555..17709
 FT /*tag= CX
 FT /number= Intron_50
 FT 17710..17864
 FT /*tag= CY
 FT /number= Exon_51
 FT 17865..18019
 FT /*tag= CZ
 FT /number= Intron_51
 FT 18020..18174
 FT /*tag= DA
 FT /number= Exon_52
 FT 18175..18329
 FT /*tag= DB
 FT /number= Intron_52
 FT 18330..18484
 FT /*tag= DC
 FT /number= Exon_53
 FT 18485..18639
 FT /*tag= DD
 FT /number= Intron_53
 FT 18640..18794
 FT /*tag= DE
 FT /number= Exon_54
 FT 18795..18949
 FT /*tag= DF
 FT /number= Intron_54
 FT 18950..19104
 FT /*tag= DG
 FT /number= Exon_55
 FT 19105..19259
 FT /*tag= DH
 FT /number= Intron_55
 FT 19260..19414
 FT /*tag= DI
 FT /number= Exon_56
 FT 19415..19569
 FT /*tag= DJ
 FT /number= Intron_56
 FT 19570..19724
 FT /*tag= DK
 FT /number= Exon_57
 FT 19725..19879
 FT /*tag= DL
 FT /number= Intron_57
 FT 19880..20034
 FT /*tag= DM
 FT /number= Exon_58
 FT 20035..20189
 FT /*tag= DN
 FT /number= Intron_58
 FT 20190..20344
 FT /*tag= DO
 FT /number= Exon_59
 FT 20345..20499
 FT /*tag= DP
 FT /number= Intron_59
 FT 20500..20654
 FT /*tag= DQ
 FT /number= Exon_60
 FT 20655..20809
 FT /*tag= DR
 FT /number= Intron_60
 FT 20810..20964
 FT /*tag= DS
 FT /number= Exon_61
 FT 20965..21119
 FT /*tag= DT
 FT /number= Intron_61
 FT 21120..21274
 FT /*tag= DU
 FT /number= Exon_62
 FT 21275..21429
 FT /*tag= DV
 FT /number= Intron_62
 FT 21430..21584
 FT /*tag= DW
 FT /number= Exon_63
 FT 21585..21739
 FT /*tag= DX
 FT /number= Intron_63
 FT 21740..21894
 FT /*tag= DY
 FT /number= Exon_64
 FT 21895..22049
 FT /*tag= DZ
 FT /number= Intron_64
 FT 22050..22204
 FT /*tag= EA
 FT /number= Exon_65
 FT 22205..22359
 FT /*tag= EB
 FT /number= Intron_65
 FT 22360..22514
 FT /*tag= EC
 FT /number= Exon_66
 FT 22515..22669
 FT /*tag= ED
 FT /number= Intron_66
 FT 22670..22824
 FT /*tag= EE
 FT /number= Exon_67
 FT 22825..22979
 FT /*tag= EF
 FT /number= Intron_67
 FT 22980..23134
 FT /*tag= EG
 FT /number= Exon_68
 FT 23135..23289
 FT /*tag= EH
 FT /number= Intron_68
 FT 23290..23444
 FT /*tag= EI
 FT /number= Exon_69
 FT 23445..23599
 FT /*tag= EJ
 FT /number= Intron_69
 FT 23600..23754
 FT /*tag= EK
 FT /number= Exon_70
 FT 23755..23909
 FT /*tag= EL
 FT /number= Intron_70
 FT 23910..24064
 FT /*tag= EM
 FT /number= Exon_71
 FT 24065..24219
 FT /*tag= EN
 FT /number= Intron_71
 FT 24220..24374
 FT /*tag= EO
 FT /number= Exon_72
 FT 24375..24529
 FT /*tag= EP
 FT /number= Intron_72
 FT 24530..24684
 FT /*tag= EQ
 FT /number= Exon_73
 FT 24685..24839
 FT /*tag= ER
 FT /number= Intron_73
 FT 24840..24994
 FT /*tag= ES
 FT /number= Exon_74
 FT 24995..25149
 FT /*tag= ET
 FT /number= Intron_74
 FT 25150..25304
 FT /*tag= EU
 FT /number= Exon_75
 FT 25305..25459
 FT /*tag= EV
 FT /number= Intron_75
 FT 25460..25614
 FT /*tag= EW
 FT /number= Exon_76
 FT 25615..25769
 FT /*tag= EX
 FT /number= Intron_76
 FT 25770..25924
 FT /*tag= EY
 FT /number= Exon_77
 FT 25925..26079
 FT /*tag= EZ
 FT /number= Intron_77
 FT 26080..26234
 FT /*tag= FA
 FT /number= Exon_78
 FT 26235..26389
 FT /*tag= FB
 FT /number= Intron_78
 FT 26390..26544
 FT /*tag= FC
 FT /number= Exon_79
 FT 26545..26699
 FT /*tag= FD
 FT /number= Intron_79
 FT 26700..26854
 FT /*tag= FE
 FT /number= Exon_80
 FT 26855..27009
 FT /*tag= FF
 FT /number= Intron_80
 FT 27010..27164
 FT /*tag= FG
 FT /number= Exon_81
 FT 27165..27319
 FT /*tag= FH
 FT /number= Intron_81
 FT 27320..27474
 FT /*tag= FI
 FT /number= Exon_82
 FT 27475..27629
 FT /*tag= FJ
 FT /number= Intron_82
 FT 27630..27784
 FT /*tag= FK
 FT /number= Exon_83
 FT 27785..27939
 FT /*tag= FL
 FT /number= Intron_83
 FT 27940..28094
 FT /*tag= FM
 FT /number= Exon_84
 FT 28095..28249
 FT /*tag= FN
 FT /number= Intron_84
 FT 28250..28404
 FT /*tag= FO
 FT /number= Exon_85
 FT 28405..28559
 FT /*tag= FP
 FT /number= Intron_85
 FT 28560..28714
 FT /*tag= FQ
 FT /number= Exon_86
 FT 28715..28869
 FT /*tag= FR
 FT /number= Intron_86
 FT 28870..29024
 FT /*tag= FS
 FT /number= Exon_87
 FT 29025..29179
 FT /*tag= FT
 FT /number= Intron_87
 FT 29180..29334
 FT /*tag= FU
 FT /number= Exon_88
 FT 29335..29489
 FT /*tag= FV
 FT /number= Intron_88
 FT 29490..29644
 FT /*tag= FW
 FT /number= Exon_89
 FT 29645..29799
 FT /*tag= FX
 FT /number= Intron_89
 FT 29800..29954
 FT /*tag= FY
 FT /number= Exon_90
 FT 29955..30109
 FT /*tag= FZ
 FT /number= Intron_90
 FT 30110..30264
 FT /*tag= GA
 FT /number= Exon_91
 FT 30265..30419
 FT /*tag= GB
 FT /number= Intron_91
 FT 30420..30574
 FT /*tag= GC
 FT /number= Exon_92
 FT 30575..30729
 FT /*tag= GD
 FT /number= Intron_92
 FT 30730..30884
 FT /*tag= GE
 FT /number= Exon_93
 FT 30885..31039
 FT /*tag= GF
 FT /number= Intron_93
 FT 31040..31194
 FT /*tag= GG
 FT /number= Exon_94
 FT 31195..31349
 FT /*tag= GH
 FT /number= Intron_94
 FT 31350..31504
 FT /*tag= GI
 FT /number= Exon_95
 FT 31505..31659
 FT /*tag= GJ
 FT /number= Intron_95
 FT 31660..31814
 FT /*tag= GK
 FT /number= Exon_96
 FT 31815..31969
 FT /*tag= GL
 FT /number= Intron_96
 FT 31970..32124
 FT /*tag= GM
 FT /number= Exon_97
 FT 32125..32279
 FT /*tag= GN
 FT /number= Intron_97
 FT 32280..32434
 FT /*tag= GO
 FT /number= Exon_98
 FT 32435..32589
 FT /*tag= GP
 FT /number= Intron_98
 FT 32590..32744
 FT /*tag= GQ
 FT /number= Exon_99
 FT 32745..32899
 FT /*tag= GR
 FT /number= Intron_99
 FT 32900..33054
 FT /*tag= GS
 FT /number= Exon_100
 FT 33055..33209
 FT /*tag= GT
 FT /number= Intron_100
 FT 33210..33364
 FT /*tag= GU
 FT /number= Exon_101
 FT 33365..33519
 FT /*tag= GV
 FT /number= Intron_101
 FT 33520..33674
 FT /*tag= GW
 FT /number= Exon_102
 FT 33675..33829
 FT /*tag= GX
 FT /number= Intron_102
 FT 33830..33984
 FT /*tag= GY
 FT /number= Exon_103
 FT 33985..34139
 FT /*tag= GZ
 FT /number= Intron_103
 FT 34140..34294
 FT /*tag= HA
 FT /number= Exon_104
 FT 34295..34449
 FT /*tag= HB
 FT /number= Intron_104
 FT 34450..34604
 FT /*tag= HC
 FT /number= Exon_105
 FT 34605..34759
 FT /*tag= HD
 FT /number= Intron_105
 FT 34760..34914
 FT /*tag= HE
 FT /number= Exon_106
 FT 34915..35069
 FT /*tag= HF
 FT /number= Intron_106
 FT 35070..35224
 FT /*tag= HG
 FT /number= Exon_107
 FT 35225..35379
 FT /*tag= HH
 FT /number= Intron_107
 FT 35380..35534
 FT /*tag= HI
 FT /number= Exon_108
 FT 35535..35689
 FT /*tag= HJ
 FT /number= Intron_108
 FT 35690..35844
 FT /*tag= HK
 FT /number= Exon_109
 FT 35845..35999
 FT /*tag= HL
 FT /number= Intron_109
 FT 36000..36154
 FT /*tag= HM
 FT /number= Exon_110
 FT 36155..36309
 FT /*tag= HN
 FT /number= Intron_110
 FT 36310..36464
 FT /*tag= HO
 FT /number= Exon_111
 FT 36465..36619
 FT /*tag= HP
 FT /number= Intron_111
 FT 36620..36774
 FT /*tag= HQ
 FT /number= Exon_112
 FT 36775..36929
 FT /*tag= HR
 FT /number= Intron_112
 FT 36930..37084
 FT /*tag= HS
 FT /number= Exon_113
 FT 37085..37239
 FT /*tag= HT
 FT /number= Intron_113
 FT 37240..37394
 FT /*tag= HU
 FT /number= Exon_114
 FT 37395..37549
 FT /*tag= HV
 FT /number= Intron_114
 FT 37550..37704
 FT /*tag= HW
 FT /number= Exon_115
 FT 37705..37859
 FT /*tag= HX
 FT /number= Intron_115
 FT 37860..38014
 FT /*tag= HY
 FT /number= Exon_116
 FT 38015..38169
 FT /*tag= HZ
 FT /number= Intron_116
 FT 38170..38324
 FT /*tag= IA
 FT /number= Exon_117
 FT 38325..38479
 FT /*tag= IB
 FT /number= Intron_117
 FT 38480..38634
 FT /*tag= IC
 FT /number= Exon_118
 FT 38635..38789
 FT /*tag= ID
 FT /number= Intron_118
 FT 38790..38944
 FT /*tag= IE
 FT /number= Exon_119
 FT 38945..39099
 FT /*tag= IF
 FT /number= Intron_119
 FT 39100..39254
 FT /*tag= IG
 FT /number= Exon_120
 FT 39255..39409
 FT /*tag= IH
 FT /number= Intron_120
 FT 39410..39564
 FT /*tag= II
 FT /number= Exon_121
 FT 39565..39719
 FT /*tag= IJ
 FT /number= Intron_121
 FT 39720..39874
 FT /*tag= IK
 FT /number= Exon_122
 FT 39875..40029
 FT /*tag= IL
 FT /number= Intron_122
 FT 40030..40184
 FT /*tag= IM
 FT /number= Exon_123
 FT 40185..40339
 FT /*tag= IN
 FT /number= Intron_123
 FT 40340..40494
 FT /*tag= IO
 FT /number= Exon_124
 FT 40495..40649
 FT /*tag= IP
 FT /number= Intron_124
 FT 40650..40804
 FT /*tag= IQ
 FT /number= Exon_125
 FT 40805..40959
 FT /*tag= IR
 FT /number= Intron_125
 FT

PD 23-DEC-1993.
 PF 14-JUN-1993: U05701.
 PR 12-JUN-1992: US-897788.
 PR 20-NOV-1992: US-979638.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI: 94-007542/01.
 DR P-PSDB: R53287.
 PT Isolated C. elegans cell death genes ced-3 and ced-4 - used to
 PS develop agents to increase or prevent cell death in organisms
 PS Claim 14, Fig 4: 127pp. English.
 CC The sequences given in Q64735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;

Best Local Similarity 99.9%; Pred. No. 0.00e+00; Indels 0; Gaps 0;

Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 3601 tcatgtgttctgcatatgaatgacacaaatagacaaattttatataatttccctat 3660
 |||||
 QY 3601 TCATGTGTGTCATATGATGACGAAATATACAAATTTTATATATTTCCCTAT 3660
 |||||
 DB 3661 tcatgtgtcgaagaaatagtaaaagcgatcatcttttcgcacatttttaccatga 3720
 |||||
 QY 3661 TCATGTGTGTCGAAATATGTAATAAAGCGCATCTTTTCGACATTTTTCATGCA 3720
 |||||
 DB 3721 acgacagctcattcatcatgctgtaagacgagacgagcgagagaataaccacacatttct 3780
 |||||
 QY 3721 ACGACAGCTCATCTTACATGCTGAAGACGAGACGCGGAGAAATACACACATCTTCT 3780
 |||||
 DB 3781 ggcgtctcgtctcagcatgtaaatgggacatcgctcgatgtaaaaaatgctgaata 3840
 |||||
 QY 3781 GCGTCTCTGCTTCAGCATGTGAATGGGATCGTCGATGTAAATAATGTGGAATA 3840
 |||||
 DB 3841 atgtaaaaatgcagcgcttttttaccatttctgcacaaatgataatgggggaaatgct 3900
 |||||
 QY 3841 ATGTAAAAATGCATGCGCTTTTTCACCTTTTCGCAAAATGTAATGGGGGAAATGT 3900
 |||||
 DB 3901 attaaaaatcatttttctgatttttcaacatcacatgatacccatatttttctgct 3960
 |||||
 QY 3901 ATTAAAAATCATTTTGTATTTTTCACATCATGATTAACCCCATATTATTTCGCT 3960
 |||||
 DB 3961 gagcaactaaaaagtagaataatagagcgaaaaaccaaatttctcaagatatacc 4020
 |||||
 QY 3961 GAGCAACTTAAAAAGTAGAATATATAGAGCAAAAACCAAAATTTCTCAAGATATTACC 4020
 |||||
 DB 4021 ttatgtgaataatgatagatgtaataagcatatcttgaatgaaatgcgcaaaaatgct 4080
 |||||
 QY 4021 TTTATGATTAATTAATGATGTTAATTAAGCATATCTTGAATGAAGTCAAGCAAAATATGT 4080
 |||||
 DB 4081 ggcgaacaccctgaaaaaaatcaaaaatctgcgaaaaatgaaaaaatgatatataca 4140
 |||||
 QY 4081 GCGAAACACCTGAAAAAATCAAAATTTCTGGGAAATTTGAAAAAATGATTAATAATACA 4140

DB 4141 ttgttcatttttctacatcatgaatgtagaaatataaaggagaaatcaaaatttcta 4200
 |||||
 QY 4141 TTTTTCATTTTTCATCATCATCATCATGATAGTAAATTAAGGAAATCAAAATTTCTA 4200
 |||||
 DB 4201 gaggataataatgtaataaacaatttcgaaatataaattgtcgaaaacgccaagaaga 4260
 |||||
 QY 4201 GAGGATATTAATGATGAAGAACATTCGGAATTAATGTGCGAAACGTCAAAAAGAGGA 4260
 |||||
 DB 4261 aatttggatcaaaaatgatcatcttaaaacacacatttcagctcgcaactttcat 4320
 |||||
 QY 4261 AATTTGGATCAAAATTCATCTCTTAACCAACATTTTCAGATCCGCCAACTTTTAT 4320
 |||||
 DB 4321 tcaccgagatcttcttcctgcgatacagttcaagtcgaatcgtcatcattcagcaagct 4380
 |||||
 QY 4321 TCACCGAGATCTTCTCTCGGATAGTCAAGTCGTAATGCTCATTCAGCAAAAGCTT 4380
 |||||
 DB 4381 ctggaccacactcaatacatatccatgtaagagagatgaactttgtgatgaccaaaca 4440
 |||||
 QY 4381 CTGGACCAACTCAATACATATTCATGAGAGAGATATGAACCTTCTCGATGCACCAACCA 4440
 |||||
 DB 4441 taagccgttttcgacgagaaacacatgacagaacttcgagtcctcgtggaaatgt 4500
 |||||
 QY 4441 TAAGCGTGTTTTCGACGAGAAACCATGTACAGAACTTCGAGTCTCGTGGAAATGT 4500
 |||||
 DB 4501 gctcatcataataatgaacacttgagcagatgcacacaggaatggtaccagagccg 4560
 |||||
 QY 4501 GCTCATCATTAATTAATGAACACTTTGACGACAGATGCCAACAGGAATGTGTACCAAGCGCG 4560
 |||||
 DB 4561 acaagacacatcttaccatattgttcagatgcatggcta 4600
 |||||
 QY 4561 ACAAGSACATCTTACCAATTTGTTGATGATGGGCTA 4600

RESULT 7
 ID 064739 standard; DNA: 7653 BP.
 AC 064739;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (56297A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key
 FT repeat_unit Location/Qualifiers
 FT 1356..1472
 FT /*tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1490-1614"
 FT 1490..1614
 FT /*tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /*tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /*tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /*tag= e
 FT /number= Exon_2
 FT 2576..2853
 FT /*tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /*tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /*tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /*tag= i

QY 3961 GAGCACTTAAAAAGTAGAGATATTAGAGGAAACCAAAATTTCTCAAGATATTACC 4020
 Db 4021 ttcttataataatagatgtaataagcatabcttgaatgaagtgtaagcaaaatattgt 4080
 QY 4021 TTTTATGATATATTAGATGTTAATAAGCATATCTTGAATGAAGTACGCAAAAATATGT 4080
 Db 4081 gccgaacacccggaaaaaatcaaaaattctcgsgaaaaattgaaaaaatcattaaatata 4140
 QY 4081 GCGAACAACCTGAAAAAATCAAAAATTTGCGAAAAATTGAAAAAATGCAATTAATAATACA 4140
 Db 4141 ttcttgatctttctacatcacatgaaatgtagaaaattaaaggaagaaatcaaaatttcta 4200
 QY 4141 TTTTTCATTTTCTACATCAGATGATGAGAAAATTAAAGGAATCAAAATTTCTTA 4200
 Db 4201 gaggaataaattgaatgaacaattcggaattaaatgctgcgaacgtcaaaaaagagga 4260
 QY 4201 GAGGATATATATGTAATGAATGAACATGCGAAAAATTAAATGTGCGAAACGCTCAAAAAAGAGA 4260
 Db 4261 aatttggtatcaaatccgattccctaaacacacacatttaagatccgcaactcttcac 4320
 QY 4261 AATTGGGTATCAAAATCGATCCTAAAACCAACACATTTCAAGATCCGCCAACTCTTCAT 4320
 Db 4321 tcaccggaatgctctctcctcgatcacatgtaacgtcgaatcgtcattcagcaagcctt 4380
 QY 4321 TCACCGGATGCTTCTCTCGGATACAGTTCAAGTGTGTAATGCTCATTTACGCAAAAGCTT 4380
 Db 4381 ctggaccacaactcaatcatattccatgaagaagatatagaacttctgcgatgcacaaaca 4440
 QY 4381 CTGACACCACTCAATATCATATTCATGAAAGATGAAAGATTTGCGATGCACCAACCA 4440
 Db 4441 taaagcgtgttttcgacgagaaaaacacatgacgaacattctgaatccctcggaatgt 4500
 QY 4441 TAAAGCCTGTGTTTCGACGAGAAAACCATGTACAGAAACTTCTGAGTCCCTGTGGAATGT 4500
 Db 4501 gccctacataaataatgaacactttgaagcagatgccaacaggaatgtaaccaagccg 4560
 QY 4501 GCCTCATCATTAATTAATGAAACACTTTGAGCAGATGCCAAACGGAATGTACCAAGCCG 4560
 Db 4561 acaagagacaatcttaccatttctcagatgcatgggcta 4600
 QY 4561 ACAAGGACAATCTTACCAATTTGTCAGATGCATGGGCTA 4600

RESULT 6

ID 064744 standard; DNA; 7653 BP.
 AC 064744;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (G6536A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophobic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key
 FT repeat_unit
 FT 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /tag= e
 FT /number= Exon_2
 FT 2576..2653
 FT /tag= f

FT /number= Intron_2
 FT 2834..3107
 FT /tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= m
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= n
 FT /number= Intron_4
 FT 4688..4719
 FT /tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5347..5760
 FT /tag= q
 FT /number= Exon_5
 FT 5761..5814
 FT /tag= r
 FT /number= Intron_5
 FT 5815..5942
 FT /tag= s
 FT /number= Exon_6
 FT 5943..6297
 FT /tag= t
 FT /number= Intron_6
 FT 6067..6138
 FT /tag= u
 FT /rpt_type= INVERTED
 FT 6298..6537
 FT /tag= v
 FT /number= Exon_7
 FT 6538..7012
 FT /tag= w
 FT /note= "G>A, from allele n2426"
 FT 6538..7012
 FT /tag= x
 FT /number= Intron_7
 FT 6567..6625
 FT /tag= y
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /tag= z
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7012..7075
 FT /tag= aa
 FT /number= Exon_8

PN W09325685-A.

FT	/number= Exon_1
FT	2367..2429
FT	/*tag= d
FT	/number= Intron_1
FT	2450..2575
FT	/*tag= e
FT	/number= Exon_2
FT	2576..2853
FT	/*tag= f
FT	/number= Intron_2
FT	2854..3107
FT	/*tag= g
FT	/number= Exon_3
FT	3108..4302
FT	/*tag= h
FT	/number= Intron_3
FT	3126..3243
FT	/*tag= i
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3329-3396"
FT	3329..3396
FT	/*tag= j
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3126-3243"
FT	3487..3759
FT	/*tag= k
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3782-4070"
FT	3782..4070
FT	/*tag= l
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 3487-3759"
FT	4303..4634
FT	/*tag= m
FT	/number= Exon_4
FT	4635..5546
FT	/*tag= n
FT	/number= Intron_4
FT	4688..4719
FT	/*tag= o
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5221..5330
FT	/*tag= p
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5547..5760
FT	/*tag= q
FT	/number= Exon_5
FT	5761..5814
FT	/*tag= r
FT	/number= Intron_5
FT	5815..5942
FT	/*tag= s
FT	/number= Exon_6
FT	5943..6297
FT	/*tag= t
FT	/number= Intron_6
FT	6062..6138
FT	/*tag= u
FT	/rpt_type= INVERTED
FT	6298..6537
FT	/*tag= v
FT	/number= Exon_7
FT	6372
FT	/*tag= w
FT	/note= "G>A, from allele n1286"
FT	6538..7012
FT	/*tag= x
FT	/number= Intron_7
FT	6567..6625
FT	/*tag= y
FT	/rpt_type= INVERTED

Query Match	99.8%	Score 998	DB 9	Length 7653
Best Local Similarity	99.9%	Pred. No. 0.00e+00		
Matches	999	Conservative	0	Mismatches 1
3601	tcattgtgttgcataatgcaatgcaaaaatagcaatttttatatattccctat	3660		
3601	tcattgtgttgcataatgcaatgcaaaaatagcaatttttatatattccctat	3660		
3661	tcattgtgtgcgaataatagcaaaaatagcaatgcaatttttgcacatg	3720		
3661	tcattgtgtgcgaataatagcaaaaatagcaatgcaatttttgcacatg	3720		
3721	acgacagcctcattccatcgtcgaagacgaagacgcgcgaataatgcaatg	3780		
3721	acgacagcctcattccatcgtcgaagacgaagacgcgcgaataatgcaatg	3780		
3781	gcgctctcgtcctcgaatgcaatgcaatgcaatgcaatgcaatgcaatg	3840		
3781	gcgctctcgtcctcgaatgcaatgcaatgcaatgcaatgcaatgcaatg	3840		
3841	attgtaaaatgcacgcgttttttaccatttgcgcaaaatgataaggggaaatgt	3900		
3841	attgtaaaatgcacgcgttttttaccatttgcgcaaaatgataaggggaaatgt	3900		
3901	atgtaaaatgcacgcgttttttaccatttgcgcaaaatgataaggggaaatgt	3960		
3901	atgtaaaatgcacgcgttttttaccatttgcgcaaaatgataaggggaaatgt	3960		
3961	gagcaacttaaaagtaggaattatgagcgaataacccaatcttcaagatatcc	4020		
3961	gagcaacttaaaagtaggaattatgagcgaataacccaatcttcaagatatcc	4020		

```

FT      exon      6298..6537
FT      /tag= w
FT      /number= Exon_7
FT      intron    6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit 6567..6625
FT      /tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      Repeat_unit 6905..6965
FT      /tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      Exon      7012..7075
FT      /tag= aa
FT      /number= Exon_8
FT      W09325685-A.
FT      23-DEC-1993.
FT      14-JUN-1993: U05701.
FT      12-JUN-1992: US-897788.
FT      20-NOV-1992: US-979638.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Horvitz HR, Shahan S, Yuan J;
FT      WPI; 94-007542/01.
FT      P-PSDB: R53282.
FT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT      develop agents to increase or prevent cell death in organisms
FT      PS Claim 14; Fig 4; 12pp; English.
FT      The sequences given in 064735-45 represent mutations of the C. elegans
FT      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT      was most abundant in embryos, but was also detected in larvae and young
FT      adults, suggesting that ced-3 is expressed not only in cells undergoing
FT      cell death. The four largest introns as well as sequences 5' of the
FT      start codon contain repetitive elements, some of which have been
FT      characterised in non-coding regions of other C. elegans genes, such
FT      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT      length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT      region that might be a transmembrane region. One region of Ced-3 is
FT      very rich in serine. It is thought that this region is involved in
FT      protein-protein interactions, similar to acid blobs in transcription
FT      factors. Of the mutations which occur within the ced-3 gene, eight of
FT      the mutations are missense mutations, two are nonsense mutations and
FT      two are splicing mutations. These mutations establish the
FT      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT      function is not essential for viability. The ced-3 and ced-4 gene
FT      products may be used to develop agents for treating conditions
FT      characterised by cell deaths, such as myocardial infarction, stroke,
FT      degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT      CC infection, aging or hair loss.
FT      CC Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
SQ
Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      3841 ATGTAAAAATGATGGTGTGTTTACACTTCTCTCCACAAATGATAGGGGAAATGCT
DB      3901 attaaatcatatttttgatattttcaacatcacatgataaccatatttttcgt
QY      3901 ATTAAATACATTTTGTGTTTTCACATCCATGATTTAACCCATTTTTCGTT
DB      3961 gagcaacttaaaaagtagaatalatagagcgaaacccaattttctcaagatalacc
QY      3961 GAGCAACTTAAAAAGTAGAATATTAGAGCGGAAACCAAAATTTCTTCAAGTATTACC
DB      4021 ttattgatataatagatgtttaataagcaatcttgatgaatgcagcaaaatattg
QY      4021 TTTATTGATATTATTAGATGTTTATAGCATATCTTGAATGAAAGTCACCAAAATATGT
DB      4081 gcaaacaccctgaaaaaatacaaaaattctgcgaaaaattgaaaaatgattaaataca
QY      4081 GCGAACACCTGAAAAAATCAAAAATTTCTGCCAAATTTGAAAAATGCAATTAATACA
DB      4141 ttttgcaattttctacatcacatgaatglaaaaaatlaaagggaatcaaatctta
QY      4141 TTTTGCAATTTTCTACATCAGATGATGATAAATTTAAAGGGAATCAAAATTTCTA
DB      4201 gagatataattgaatgaacaattgcgaattaaatgtgcgaacgtcaaaaagaga
QY      4201 GAGGATATATTGATGAAACATTTGCGAAATTTAAATGTGCGAAACGTCAAAAAGAGA
DB      4261 aattggatcaacaatgcataccttaaacacacacatttcagcatccgcaacttcgt
QY      4261 AATTGGGATTCAAATTCATCTTAACCAACACATTTTCACATTCGCCAATCTTCAT
DB      4321 tcaacgagatctctctctcgcgatacagttcaagttcaatcgtcaatcagcaagctt
QY      4321 TCACCGAGATGCTTCTCTCGATACAGTTCAAGTCGTAATGTCATTCACCAAAAGCT
DB      4381 ctgagcaactcaatatacatatccatgaagagatgaacttttcgatcaccaacca
QY      4381 CTGAGCAACTCAATATACATATTCATGAAAGAGATGAACTTTCATGATGACCAACCA
DB      4441 taagccgtgttttcgcgcgaaacacatgtacagaacttcgcggtctcgtggaatgt
QY      4441 TAAGCCGTGTTTTCGACGGAAGAACCATGTACAGAAATCTTCGAGTCTCGTGAATGT
DB      4501 gctcatcaataaatgacaactttgagcagatgccaacgcgaatgtlaaccaagcgt
QY      4501 GCCTCATATAATATATGAACTTTGAGCAGATGCCAACGGAATGTAACCAAGCGC
DB      4561 acaaggacaacttaccaaattgttcagatgcagtgcta 4600
QY      4561 ACAAGACAATCTTACCAATTTGTTCAGATGATGGGCTA 4600

```

```

RESULT 5
ID 064741 standard; DNA: 7653 BP.
AC 064741.
DE 23-JUN-1994 (first entry)
KW ced-3 (G6372A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c

```

Db 3721 acgacagctcacttcacatgctgaagacgagagacgaggaataataccacacatcttct 3780
 |||
 Qy 3721 ACACAGCTCCTCACTTCACATGCTGAAGACGAGACGCGGAGAAATACACACATCTTCT 3780
 Db 3781 ggcctcctcgtcttcagcatgtaaatggatcctcgctcgtatgaataaaatgctgaata 3840
 |||
 Qy 3781 GCGTCTCTGCTCTTCAGCATGTCAAAATGGATCTCGTGTGATGTAAAAAATCTCGAATA 3840
 Db 3841 atgtaaaaaatgcatgcttcttttctacacttctctgcacaaatgaataggggaaatgt 3900
 |||
 Qy 3841 ATGTAAAAATGATCGCTGTTTTTTTACCTTTCTTCACAAATGATAGGGGAAATGT 3900
 Db 3901 attaaaataatcttctgtatcttctcaacatcagatgaatacccatatttctgt 3960
 |||
 Qy 3901 ATTAAATATCATTTTGTGATTTTTCACATCAACATGATTAACCCCATTTTTCGT 3960
 Db 3961 gagcaacttaaaagttagaataatttagagcgaaaccaaacttctcaagatattacc 4020
 |||
 Qy 3961 GAGCAACTTAAAAAGTAGAATATTAGACGAAACCAAAATTTCTTCAAGATATTAC 4020
 Db 4021 ttattgataataatagatgtaataaagacatcttgaatgaatgcaagcaaaatagt 4080
 |||
 Qy 4021 TTTATTGATATTATAGATGTATTATAGCATATCTGAATGAAGTCAACCAAAATATGT 4080
 Db 4081 gggaaacacctgaaaaaatcaaaatctcggaataattgaaaaatgcatataataa 4140
 |||
 Qy 4081 GCGAAACACCTGAAAAAATCAAAAAATCTGCAAAATTTGAAAAATGATTAATATA 4140
 Db 4141 ttttgcattcttcacatcacatgtaagtagaanaatgaaggaacaaatcttca 4200
 |||
 Qy 4141 TTTTGCATTTTCTCAATCACATGAATGAGAAATTTAAAGGGAATCAAAATTTCTA 4200
 Db 4201 gaggataataatgaatgaacacatgtagaataatgaatgtagaagcaacgtcaaaaagaga 4260
 |||
 Qy 4201 GAGGATTAATTAATGAATGAACATTTGCAAAATTTAAATGTGCGAAAGTCAAAAGAGAGA 4260
 Db 4261 aattggtatcaaaatcgatccttaaaacacacattcaagatcgcgaacttcat 4320
 |||
 Qy 4261 AATTGGGTATCAAAATCGATCTTAAACCAACATTTTCAGATCTCGCAACTCTTCT 4320
 Db 4321 tcaacgagatctctctcgcgatacaagtcgaatgtaactgcctcattcagcaaacct 4380
 |||
 Qy 4321 TCACCGGATGCTCTCTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 4380
 Db 4381 ctggacaactcaatcacatcttcagagaagagatgaacttctcgtatgacacaca 4440
 |||
 Qy 4381 CTGGACCACTCAATCATATCTCATGTAAGAGATATGAACITTTGATGACACCAACA 4440
 Db 4441 taagcgtgttttcgagagaagaacacatgtagaagaactctcgtatcctcgtggaatgt 4500
 |||
 Qy 4441 TAACTGCTGTTTTCGAGAGAAAACATGTACAGAACTTCTGAGTCTCTGAGATGT 4500
 Db 4501 gccctacataaataatgaacactttagagagatgcaacacaggaatggtaccagaagcgt 4560
 |||
 Qy 4501 GCCCTACATAAATAATGAACACTTTGAGCAGATGCCAACACGGAATGATCAAGGCCG 4560
 Db 4561 acaaggaacatcttaccatattgttcagatgcatggccta 4600
 |||
 Qy 4561 ACAAGGACAAATCTTACCAATTTGTCAGATGATGGGCTA 4600

RESULT 4

ID 064738 standard: DNA; 7653 BP.
 AC 064738;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (C5940T) gene.
 KW C. elegans; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FT key location/Qualifiers
 FT repeat_unit 1356..1472

FT FT /*tag= a
 FT FT /rpl_type= INVERTED
 FT FT /note= "Inverted w.r.t. repeat at 1490-1614"
 FT FT 1490..1614
 FT FT /*tag= b
 FT FT /rpl_type= INVERTED
 FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT FT 2232..2366
 FT FT /*tag= c
 FT FT /number= Exon_1
 FT FT 2367..2429
 FT FT /*tag= d
 FT FT /number= Intron_1
 FT FT 2450..2575
 FT FT /*tag= e
 FT FT /number= Exon_2
 FT FT 2576..2853
 FT FT /*tag= f
 FT FT /number= Intron_2
 FT FT 2854..3107
 FT FT /*tag= g
 FT FT /number= Exon_3
 FT FT 3108..4302
 FT FT /*tag= h
 FT FT /number= Intron_3
 FT FT 3126..3243
 FT FT /*tag= i
 FT FT /rpl_type= INVERTED
 FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT FT 3329..3396
 FT FT /*tag= j
 FT FT /rpl_type= INVERTED
 FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT FT 3487..3759
 FT FT /*tag= k
 FT FT /rpl_type= INVERTED
 FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT FT 3782..4070
 FT FT /*tag= l
 FT FT /rpl_type= INVERTED
 FT FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT FT 4303..4634
 FT FT /*tag= m
 FT FT /number= Exon_4
 FT FT 4635..5546
 FT FT /*tag= n
 FT FT /number= Intron_4
 FT FT 4688..4719
 FT FT /*tag= o
 FT FT /rpl_type= INVERTED
 FT FT /note= "Repeat 3"
 FT FT 5221..5330
 FT FT /*tag= p
 FT FT /rpl_type= INVERTED
 FT FT /note= "Repeat 3"
 FT FT 5547..5760
 FT FT /*tag= q
 FT FT /number= Exon_5
 FT FT 5761..5814
 FT FT /*tag= r
 FT FT /number= Intron_5
 FT FT 5815..5942
 FT FT /*tag= s
 FT FT /number= Exon_6
 FT FT 5957..6138
 FT FT /*tag= t
 FT FT /note= "C>T, from allele n1165"
 FT FT 5943..6297
 FT FT /*tag= u
 FT FT /number= Intron_6
 FT FT 6062..6138
 FT FT /*tag= v
 FT FT /rpl_type= INVERTED

23-JUN-1994 (first entry)
 DE ced-3 (G2487A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 PH Key location/Qualifiers
 FT repeat_unit 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /tag= d
 FT /number= Intron_1
 FT 2430..2575
 FT /tag= e
 FT /number= Exon_2
 FT 2487
 FT /tag= f
 FT /note= "G>A, from allele n718"
 FT 2576..2853
 FT /tag= g
 FT /number= Intron_2
 FT 2854..3107
 FT /tag= h
 FT /number= Exon_3
 FT 3108..4302
 FT /tag= i
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= m
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= n
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= o
 FT /number= Intron_4
 FT 4688..4719
 FT /tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= q
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /tag= r
 FT /number= Exon_5
 FT 5761..5814
 FT /tag= s
 FT /number= Intron_5

FT exon 5815..5942
 FT /tag= t
 FT /number= Exon_6
 FT 5943..6297
 FT /tag= u
 FT /number= Intron_6
 FT 6062..6138
 FT /tag= v
 FT /rpt_type= INVERTED
 FT 6298..6537
 FT /tag= w
 FT /number= Exon_7
 FT 6538..7012
 FT /tag= x
 FT /number= Intron_7
 FT 6567..6625
 FT /tag= y
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /tag= z
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7012..7075
 FT /tag= aa
 FT /number= Exon_8
 FT 709325685-A.
 PN 23-DEC-1993.
 PD 14-JUN-1993; U05701.
 PE 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PI (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI; 94-007542/01.
 DR P-PSDB; R53280.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 14; Fig 4; 127pp; English.
 CC The sequences given in 064735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
 Query Match 99.8%; Score 998; DB 9; Length 7653;
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;
 Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 3601 tcattgtgttgcacgaataatgacccaataatgacatttttatataattccctat 3660
 QY |||||||
 Db 3601 TCATGTGTTGTCATATGATGACGAAATATTTTATATTTTCCCTAT 3660
 QY |||||||
 Db 3661 tcattgtgtcacaataatgacccaataatgacattttcgacattttatataatga 3720
 QY |||||||
 Db 3661 TCATGTGTTGTCACAAATATGATGACGCAATGCAATTTTTCACATGCA 3720
 QY |||||||

Dp	3601	tcatggtggttgcataatgaagcaaaaatagaacttttataattttcccat	3650
Oy	3601	TCATGTGTTGTTCTCATATGAAGCAAAAATAGCAATTTTATATATTTTCCCTAT	3660
Dp	3661	tcatgttgtcagaaaatagtaaaaaagcagatgatttttcagacattttacatcga	3720
Oy	3661	TCATGTGTTGTTCTCATATGAAGCAAAAATAGCAATTTTATATATTTTCCCTAT	3730
Dp	3721	acgacagctcaatcacaagtcgaagacgagagaaatacacacattttct	3780
Oy	3661	TCATGTGTTGAGAAAATAGTAAAAAGGCATGATTTTCGACATTTTTCATACATGA	3720
Dp	3721	acgacagctcaatcacaagtcgaagacgagagaaatacacacattttct	3780
Oy	3721	ACGACAGCTCACTTCACATCTGMAACAGAGACGCGAGAAAATCCACACATCTTCT	3780
Dp	3781	gcgccttcgtcttcacagatgaagaatcgctgcgtgcgtatgtaaaaaaatgctgaata	3840
Oy	3781	GCCTCTCTCGTTCCACAGATGTAATGGGATCTCGTGCATGTAAAAAATGTGAAAT	3840
Dp	3841	atgtaaaaatgcatgagtttttttaccatttcgtcacaaatgtaataaggggaaaatggt	3900
Oy	3841	ATGTAAAAATGCATGGGTTTTCACATTTCGTGCACAAATGAAATGGGGAAAATGT	3900
Dp	3901	atataaatacaatttttgtatatttcaacatacatgatataaaccatatttttgtt	3960
Oy	3901	ATTAATAATCAATTTTGTATTTTTCACATCAACATGATTAACCCCATATTTTTCCTT	3960
Dp	3961	gagaaacttaaaagtagaagaatattagagcgaaaaacaaatttcctcaagatattcc	4020
Oy	3961	GAGCAACTTAAAAAGTAGAATATTAGCGCAAAACCAAAATTTCTTCAGATTTTACC	4020
Dp	4021	tttatgtataattatagatglttaataagatalctgtgaatgaagtcgcgcaaaaatggt	4080
Oy	4021	TTTATGTATATTATTAAGATGTATTAAGCATATCTTGATGAAGTCAAGCAAAATATAGT	4080
Dp	4081	gcgaaacacactgaaaaaatccaasaattctgcgaaaaattgaaaaaatgcattaaatca	4140
Oy	4081	GCGAAACACTGAAAAAATCAAAAATTTCTGGAAATTTGAAAAAATCATTAATAATACA	4140
Dp	4141	tttttgattttcttaacatacagaatgtagaaaaattaaagggaatactcaaatrtcta	4200
Oy	4141	TTTTGCATTTTCTACATACACTGATGTAGAAAATTTAAAGGAAATTCAAAATTTCTA	4200
Dp	4201	gagagataataitgaatgaanaacatgcgaattlaaaatgctgcgaaacgtccaanaagaga	4260
Oy	4201	GAGGATTAATTTGAATGAATGAACATTTGGGAATTTAAATGTGCAAAAGTCAAAAAAGAGA	4260
Dp	4261	aatttgggtatcaaatcgatccttaaaccaacaatttcagatcgcgcacatttcatt	4320
Oy	4261	AATTTGGGTATCAAAAATCGATCCTTAACCAACACATTTTCACATCTCCCAACTTCTCAT	4320
Dp	4321	tcacaggatgcctctctcctgcgataagctaaagtcgaatgcgcatttcagagaaagctt	4380
Oy	4321	TCACAGGATGCTCTTCTCTCGGATTAAGATTCAAGTCAGTCGATGCTCATTTCCAGCAAGCTT	4380
Dp	4381	ctggaaccaactcaatacatattccatgaagaggaatgaacttttgcagtgcaccaacca	4440
Oy	4381	CTGGACCAACATCAATACATATTCATGAAGAGGATATGAACTTTGTCATGCACCAACCA	4440
Dp	4441	taagcgtgttttcgacgagaaaaacatgtacaagaaactctcgagttcctcgttggaatgt	4500
Oy	4441	TAAACCGTGTTTTCGAGGAAAAACCATGTAAAGAAACTTTCGAGTCTCGTGTGAATGT	4500
Dp	4501	gcccatcatataaatgaacacatttgcgcagatgcgaacagagaaatgtaccagaagcgt	4560
Oy	4501	GCCCATCATTAATTAATGAACACTTTTGACCAAGATGCCAACAGGAATGTACCAAGGCGC	4560
Dp	4561	acaaggaacaatctaccacaatttgttcagatgcatgagcta 4600	
Oy	4561	ACAAGGACAATCTTACCAATTTTGTTCAGATGCATGGGCTA 4600	
RESULT	3		
ID	064736	standard: DNA: 7653 BP.	
NC	064736:		

```

|||||
QY 3601 TCATGCGTTTGTTCATATGAAATGAGCAAAAATAGCAATTTTATATTTTCCCATAT 3660
Db 3661 tcatgttgcgagaaaaaagtaaaaaagcgcatgcatcttttcgcacatttttacatcga 3720
QY 3661 TCATGTTGTGCGAGAAAAAATGTAATAAAAGCGATGATTTTTCGACATTTTTCATCGA 3720
Db 3721 acgacagcctccttcacatcgcgtgaaagcgcggagaaataaccacattctct 3780
QY 3721 ACGACAGCTCCTCTCAGATGCTGAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780
Db 3781 gcgtcctcgtcttcacagcagtaagtgatcgcgtcgtatgtaaaaaaatgtaata 3840
QY 3781 GCGTCTCTGCTCTCGCATGTAATGGATCTCGTGTGATGTAATAAAATGTGGAATA 3840
Db 3841 acgtaaaaaatgcacgcttcttttaccacttttcgcacaaatgaataggggaaaaatgt 3900
QY 3841 AAGTAAAAAATGCATCGTCTTTTTCACCTTTTCGACAAATGAATGCGGAAAAATGT 3900
Db 3901 attaaatacatcttttctatttttcaacatcacatgatacccatatttttctgt 3960
QY 3901 ATTAAATACATTTTGTATTTTTCACATCAGATTAACCCCATTTATTTTCGT 3960
Db 3961 gagcaacttaaaaagtagaataattagagcgaaaaaccaaatcttccaagatactac 4020
QY 3961 GAGCACTTAAAAAGTAGAATATTAGACCGAAACCAAAATTTCTTCAAGATATTAC 4020
Db 4021 ttattgataatataagatgatttaataagcatcttgaatgaatgcagcaaaatattgt 4080
QY 4021 TTTATGATTAATATATGATGTTATATGATATCTTGAATGAAGTCAACAAAAATATGT 4080
Db 4081 gggaaacacctgaaaaaaatcaaaaaatctggaataatgaaaaaatgataaataata 4140
QY 4081 GCGAAACACCTGAAAAAAATCAAAAATCTCGAAATTTGAAAAAATGATTAATAATA 4140
Db 4141 tttttcattttctacatcacatgagatgtagaaaatgaaggaggaatcaaatctta 4200
QY 4141 TTTTTCATTTTCTCATCATCATGATTAAGAAATTTAAAGGGAATCAAAATTTTCA 4200
Db 4201 gaggataataatgaaatgaaacattgcgaaaattaaaatgtgcgaacgtaaaaaagaga 4260
QY 4201 GAGGATATATGATGAAGAAACATTTGCGAAATTAATAATGCGAAACGTCAAAAAAGAGA 4260
Db 4261 aatttgggtataaaaatcgatcctaaaccaacatttcagcatccgcgaacttctcat 4320
QY 4261 AATTtGGTATCAAAATCGATCCTTAACCAACACATTTTCACATCCGCCAATCTTCA 4320
Db 4321 tcaacgagatgcctctctcgcgatacagttcaagtcgtaactcgtcatccagcaagct 4380
QY 4321 TCACCGGATGCTCTTCTCTCGGATACAGTTCAAGTGTATGCTCATTCAGCAAAAGCT 4380
Db 4381 ctggaccaactcaatacatattccatgaaaggagataatgaacttctgcagtacacaacca 4440
QY 4381 CTGGACCAACTCAATCATATTCATGAGAGATATGAACTTTTTCGATGCACCAACCA 4440
Db 4441 taagccggttttgcgcgagaaacccatgtacagaactcttcgagtcctcgttgtaagt 4500
QY 4441 TAAGCCGTTTTCGACGAGAAACCATGTACAGAAATCTTCGATCCTCGTGAATGT 4500
Db 4501 ggcctacataaataatgaaacattgagcagatgcaacacaggaatgtaccaaagccg 4560
QY 4501 GCGTCTCTCTTAATATGACACTTTGACGAGATGCCAACAGGAATGTACCAAGCCG 4560
Db 4561 acaagacaactcttacaatttgttcagatgcatggcta 4600
QY 4561 ACAAGACAATCTTACCATTGTTCAGATGCATGGGCTA 4600

```

RESULT 2
 ID T38196 standard; DNA: 7653 BP.
 AC T38196;
 DT 17-DEC-1996 (first entry)
 DE Nematode Ced-3 gene.

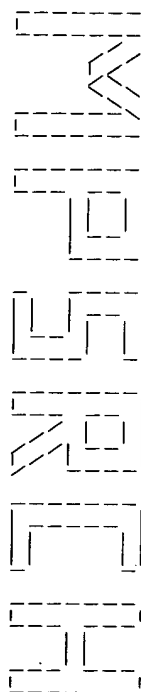
```

KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KM cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Location/Qualifiers
FT repeat_unit
  FT /tag= a
  FT /label= Repeat-1
  FT /tag= b
  FT /label= Repeat-2
  FT /tag= c
  FT /codon_start= 2232..2234
  FT /tag= d
  FT /label= T(n1040)
  FT /note= "causes L27F mutation"
  FT /tag= e
  FT /label= Intron-1
  FT /tag= f
  FT /tag= g
  FT /label= A(n718)
  FT /note= "causes G65R mutation"
  FT /tag= h
  FT /label= Intron-2
  FT /tag= i
  FT /label= Intron-3
  FT /tag= j
  FT /label= Intron-3
  FT /tag= k
  FT /label= Repeat-1
  FT /tag= l
  FT /label= Repeat-1
  FT /tag= m
  FT /label= Repeat-2
  FT /tag= n
  FT /label= Repeat-2
  FT /tag= o
  FT /label= Repeat-2
  FT /tag= p
  FT /label= Intron-4
  FT /tag= q
  FT /label= Repeat-3
  FT /tag= r
  FT /label= Repeat-3
  FT /tag= s
  FT /label= Repeat-3
  FT /tag= t
  FT /label= A(n2433)
  FT /note= "causes G360S mutation"
  FT /tag= u
  FT /label= Intron-5
  FT /tag= v
  FT /tag= w
  FT /label= T(n1165)
  FT /note= "creates premature stop codon at 403"

```

FT	/*tag= g	/note= "G>A, fromm allele n718"
FT	2576..2853	
FT	/*tag= h	
FT	/number= Intron_2	
FT	2854..3107	
FT	/*tag= i	
FT	/number= Exon_3	
FT	3108..4302	
FT	/*tag= j	
FT	/number= Intron_3	
FT	3126..3243	
FT	/*tag= k	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat at 3329-3396"	
FT	3329..3396	
FT	/*tag= l	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat at 3126-3243"	
FT	3487..3759	
FT	/*tag= m	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat at 3782-4070"	
FT	3782..4070	
FT	/*tag= n	
FT	/rpt_type= INVERTED	
FT	/note= "Inverted w.r.t. repeat 3487-3759"	
FT	4303..4634	
FT	/*tag= o	
FT	/number= Exon_4	
FT	4635..5546	
FT	/*tag= p	
FT	/number= Intron_4	
FT	4688..4719	
FT	/*tag= q	
FT	/rpt_type= INVERTED	
FT	/note= "Repeat 3"	
FT	5221..5330	
FT	/*tag= r	
FT	/rpt_type= INVERTED	
FT	/note= "Repeat 3"	
FT	5547..5760	
FT	/*tag= s	
FT	/number= Exon_5	
FT	5757	
FT	/*tag= t	
FT	/note= "G>A, from allele n2433"	
FT	5761..5814	
FT	/*tag= u	
FT	/number= Intron_5	
FT	5815..5942	
FT	/*tag= v	
FT	/number= Exon_6	
FT	5940	
FT	/*tag= w	
FT	/note= "C>T, from allele n1165"	
FT	5943..6297	
FT	/*tag= x	
FT	/number= Intron_6	
FT	6062..6138	
FT	/*tag= y	
FT	/rpt_type= INVERTED	
FT	6297	
FT	/*tag= z	
FT	/note= "G>A, fromm allele n717"	
FT	6298..6537	
FT	/*tag= aa	
FT	/number= Exon_7	
FT	6332	
FT	/*tag= ab	
FT	/note= "C>T, from allele n1949"	
FT	6372	
FT	/*tag= ac	

FT	/note= "G>A, from allele n1286"
FT	6434
FT	/*tag= ad
FT	/note= "C>T, fromm alleles n1129 and n1164"
FT	6485
FT	/*tag= ae
FT	/note= "C>T, from allele n2430"
FT	6535
FT	/*tag= af
FT	/note= "G>A, from allele n2426"
FT	6538..7012
FT	/*tag= ag
FT	/number= Intron_7
FT	6567..6625
FT	/*tag= ah
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 6905-6965"
FT	6905..6965
FT	/*tag= ai
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 6567-6625"
FT	7013..7075
FT	/*tag= aj
FT	/number= Exon_8
FT	7020
FT	/*tag= ak
FT	/note= "C>T, from allele n1163"
FT	1..5850
FT	/*tag= al
FT	/note= "Claim 2, inhibitor fragment"
FT	1..3020
FT	/*tag= am
FT	/note= "Claim 2, inhibiotr fragment"
PN	W09325694-A.
PD	23-DEC-1993.
PF	14-JUN-1993: U05705.
PR	12-JUN-1992: US-897788.
PR	20-NOV-1992: US-984182.
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.
PI	Horvitz HR, Shaham S, Yuan J;
DR	WPJ: 94-007551/01.
P-PSDB:	R45262.
PT	Agents which affect activity of cell death genes - used to develop drugs for treating conditions characterised by cell death or proliferation
PS	Disclosure: Fig 3, 132pp; English.
CC	This sequence encodes the C.elegans cell death gene, ced-3. Fragments of the amino terminal of the protein encoded by this sequence act as inhibitors of ced-3. This gene has considerable similarity to human interleukin-1beta convertase (ICE), which converts pro-interleukin-1beta to the active cytokine and is involved in inflammatory response in humans. The similarity between the two sequences suggests that inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3 proteins have an overall amino acid similarity of 28%. The ced-3 inhibitors may be used for identifying agents which affect the activity of a gene belonging to the to the ced-3/ICE family of genes and for diagnosis of diseases characterised by cell death. They can also be used to develop drugs for treating conditions characterised by cell deaths such as myocardial infarction, stroke, degenerative disease, traumatic brain injury, hypoxia, pathogenic infection, or hair loss, or drugs for reducing the proliferative capacity or size of a population of cells such as cancerous cells, cells which produce autoreactive antibodies, infected cells, hair follicle cells or cells which are critical to the life of a parasite, pest or recombinant organism. They may also be used in the diagnosis of inflammatory disease.
SQ	Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
Query Match	100.0%; Score 1000; DB 9; Length 7653;
Best Local Similarity	100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	3601 catgttgattgctcatagtaatgcagaataatggcaatttttatatatattccctcat 3660


 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Aug 7 19:20:42 1999; Maspar time 234.95 Seconds

Tabular output not generated. 911.985 Million cell updates/sec

Title: >US-08-287-669-18

Description: (3601-4600) from US08287669.seq (5 of 10)

Perfect Score: 1000

N.A. Sequence: 3601 TCATGTCGTTGTCATATG.....TTCATGATGTCATGGGCTA 4600

Comp: AGTACACCAACAAAGTATAC.....AACAGTCTACGTACCCGAT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

n.geneseq35
 1:part1 2:part3 3:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40 41:part41 42:part42 43:part43
 44:part44 45:part45 46:part46 47:part47 48:part48
 49:part49 50:part50 51:part51 52:part52 53:part53
 54:part54 55:part55 56:part56 57:part57 58:part58
 59:part59 60:part60

Statistics: Mean 8.973; Variance 6.203; scale 1.447

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1000	100.0	7653	9	054401	ced-3 gene.	0.00e+00
2	1000	100.0	7653	24	T38196	Nematode Ced-3 gene.	0.00e+00
3	998	99.8	7653	9	064738	ced-3 (G5487A) gene.	0.00e+00
4	998	99.8	7653	9	064738	ced-3 (G5487A) gene.	0.00e+00
5	998	99.8	7653	9	064741	ced-3 (G6372A) gene.	0.00e+00
6	998	99.8	7653	9	064744	ced-3 (G6536A) gene.	0.00e+00
7	998	99.8	7653	9	064739	ced-3 (G6297A) gene.	0.00e+00
8	998	99.8	7653	9	064745	ced-3 (C7020T) gene.	0.00e+00
9	998	99.8	7653	9	064742	ced-3 (G6434T) gene.	0.00e+00

10	998	99.8	7653	9	064743	ced-3 (G6485T) gene.	0.00e+00
11	998	99.8	7653	9	064740	ced-3 (G6322T) gene.	0.00e+00
12	998	99.8	7653	9	054666	ced-3 gene.	0.00e+00
13	998	99.8	7653	9	064737	ced-3 (G5757A) gene.	0.00e+00
14	298	29.8	2485	14	079970	Interleukin-1 beta co	2.66e-19
15	127	12.7	7653	9	064738	ced-3 (G5940T) gene.	2.76e-56
16	127	12.7	7653	24	T38196	Nematode Ced-3 gene.	2.76e-56
17	127	12.7	7653	9	064736	ced-3 (G2487A) gene.	2.76e-56
18	127	12.7	7653	9	064745	ced-3 (C7020T) gene.	2.76e-56
19	127	12.7	7653	9	064739	ced-3 (G6297A) gene.	2.76e-56
20	127	12.7	7653	9	064742	ced-3 (G6434T) gene.	2.76e-56
21	127	12.7	7653	9	064743	ced-3 (G6485T) gene.	2.76e-56
22	127	12.7	7653	9	064741	ced-3 (G6372A) gene.	2.76e-56
23	127	12.7	7653	9	064737	ced-3 (G5757A) gene.	2.76e-56
24	127	12.7	7653	9	054666	ced-3 gene.	2.76e-56
25	127	12.7	7653	9	064744	ced-3 (G6536A) gene.	2.76e-56
26	127	12.7	7653	9	064740	ced-3 (C6322T) gene.	2.76e-56
27	127	12.7	7653	9	054401	ced-3 gene.	2.76e-56
28	40	4.0	91	9	051746	Oligonucleotide probe	1.44e-07
29	38	3.8	204	1	N81164	Base substituted E.co	1.43e-06
30	37	3.7	91	9	051746	Oligonucleotide probe	4.44e-06
31	37	3.7	204	1	N81164	Base substituted E.co	4.44e-06
32	33	3.3	91	46	V44650	Mammalian DNA replica	3.80e-06
33	32	3.2	91	46	V44650	Mammalian DNA replica	1.13e-03
34	30	3.0	114	12	070465	Generic DNA sequence	9.59e-03
35	30	3.0	114	12	070465	Generic DNA sequence	9.59e-03
36	30	3.0	114	12	070465	Generic DNA sequence	9.59e-03
37	30	3.0	501	3	N50025	Sequence encoding new	2.77e-02
38	29	2.9	114	12	070467	Generic DNA sequence	2.77e-02
39	28	2.8	498	3	N50034	Sequence encoding new	7.77e-02
40	28	2.8	501	3	N50029	Sequence encoding new	7.77e-02
41	28	2.8	501	3	N50032	Sequence encoding new	7.77e-02
42	28	2.8	501	3	N50027	Sequence encoding new	7.77e-02
43	28	2.8	501	3	N50031	Sequence encoding new	7.77e-02
44	28	2.8	501	3	N50030	Sequence encoding new	7.77e-02
45	28	2.8	501	3	N50024	Sequence encoding new	7.77e-02

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	054401	1000	100.0	7653	9	054401	ced-3 gene.	0.00e+00
2	054401	1000	100.0	7653	24	T38196	Nematode Ced-3 gene.	0.00e+00
3	054401	998	99.8	7653	9	064738	ced-3 (G5487A) gene.	0.00e+00
4	054401	998	99.8	7653	9	064738	ced-3 (G5487A) gene.	0.00e+00
5	054401	998	99.8	7653	9	064741	ced-3 (G6372A) gene.	0.00e+00
6	054401	998	99.8	7653	9	064744	ced-3 (G6536A) gene.	0.00e+00
7	054401	998	99.8	7653	9	064739	ced-3 (G6297A) gene.	0.00e+00
8	054401	998	99.8	7653	9	064745	ced-3 (C7020T) gene.	0.00e+00
9	054401	998	99.8	7653	9	064742	ced-3 (G6434T) gene.	0.00e+00
10	054401	998	99.8	7653	9	064743	ced-3 (G6485T) gene.	0.00e+00
11	054401	998	99.8	7653	9	064740	ced-3 (G6322T) gene.	0.00e+00
12	054401	998	99.8	7653	9	054666	ced-3 gene.	0.00e+00
13	054401	998	99.8	7653	9	064737	ced-3 (G5757A) gene.	0.00e+00
14	054401	298	29.8	2485	14	079970	Interleukin-1 beta co	2.66e-19
15	054401	127	12.7	7653	9	064738	ced-3 (G5940T) gene.	2.76e-56
16	054401	127	12.7	7653	24	T38196	Nematode Ced-3 gene.	2.76e-56
17	054401	127	12.7	7653	9	064736	ced-3 (G2487A) gene.	2.76e-56
18	054401	127	12.7	7653	9	064745	ced-3 (C7020T) gene.	2.76e-56
19	054401	127	12.7	7653	9	064739	ced-3 (G6297A) gene.	2.76e-56
20	054401	127	12.7	7653	9	064742	ced-3 (G6434T) gene.	2.76e-56
21	054401	127	12.7	7653	9	064743	ced-3 (G6485T) gene.	2.76e-56
22	054401	127	12.7	7653	9	064741	ced-3 (G6372A) gene.	2.76e-56
23	054401	127	12.7	7653	9	064737	ced-3 (G5757A) gene.	2.76e-56
24	054401	127	12.7	7653	9	054666	ced-3 gene.	2.76e-56
25	054401	127	12.7	7653	9	064744	ced-3 (G6536A) gene.	2.76e-56
26	054401	127	12.7	7653	9	064740	ced-3 (C6322T) gene.	2.76e-56
27	054401	127	12.7	7653	9	054401	ced-3 gene.	2.76e-56
28	054401	40	4.0	91	9	051746	Oligonucleotide probe	1.44e-07
29	054401	38	3.8	204	1	N81164	Base substituted E.co	1.43e-06
30	054401	37	3.7	91	9	051746	Oligonucleotide probe	4.44e-06
31	054401	37	3.7	204	1	N81164	Base substituted E.co	4.44e-06
32	054401	33	3.3	91	46	V44650	Mammalian DNA replica	3.80e-06
33	054401	32	3.2	91	46	V44650	Mammalian DNA replica	1.13e-03
34	054401	30	3.0	114	12	070465	Generic DNA sequence	9.59e-03
35	054401	30	3.0	114	12	070465	Generic DNA sequence	9.59e-03
36	054401	30	3.0	114	12	070465	Generic DNA sequence	9.59e-03
37	054401	30	3.0	501	3	N50025	Sequence encoding new	2.77e-02
38	054401	29	2.9	114	12	070467	Generic DNA sequence	2.77e-02
39	054401	28	2.8	498	3	N50034	Sequence encoding new	7.77e-02
40	054401	28	2.8	501	3	N50029	Sequence encoding new	7.77e-02
41	054401	28	2.8	501	3	N50032	Sequence encoding new	7.77e-02
42	054401	28	2.8	501	3	N50027	Sequence encoding new	7.77e-02
43	054401	28	2.8	501	3	N50031	Sequence encoding new	7.77e-02
44	054401	28	2.8	501	3	N50030	Sequence encoding new	7.77e-02
45	054401	28	2.8	501	3	N50024	Sequence encoding new	7.77e-02

Db	34483	GGGCGAGAGAAATAAAAAACCCTCCCTCTGCGTCTTCGCCCTTCACATGATGAATGCAC	34542
Oy	3753	GAGCGGAGAGAAATACCACCATCTT-TCTGGCTCTCTGCTTCACCATGATGAATGGGA	3811
Db	34543	TGTCGGTCGATGTTGAAAAATGTGCAAAAAAG	34574
Oy	3812	TCTGGTCGATGTTAAAAAATGTGCAATATG	3843

RESULT	15				
LOCUS	CEZK1073	37640 bp	DNA	INV	23-NOV-1998
DEFINITION	Caenorhabditis elegans cosmid ZK1073, complete sequence.				
ACCESSION	Z68135				
NID	91753051				
VERSION	Z68135.1	GI:1753051			
KEYWORDS	HTG; Transposon.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis				
REFERENCE	1 (bases 1 to 37640)				
AUTHORS	McMurray, A.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-NOV-1995) Louis, MO 63110, USA. E-mail:				
REFERENCE	jesse@anger.ac.uk or trhematode.wustl.edu				
AUTHORS	2 (bases 1 to 37640)				
	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Larellle, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, R., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Spout, J., and Wohlman, P.				
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans				
JOURNAL	Nature 368 (6466), 32-38 (1994)				
MEDLINE	94150718				
COMMENT	On Dec 25, 1996 this sequence version replaced gi:1100866. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder /P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: http://webace.sanger.ac.uk/cqi-bin/display?db=wormacc&class=sequence&object=ZK1073				

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence is the entire insert of clone ZK1073. The true left end of clone F0968 is at 30045 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z70681. The end of this sequence (37537..37640) overlaps with the start of sequence Z68132.

```

source
1. .37640
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="ZK1073"
23268. .25951
/gene="ZK1073.2"
join(23268. .23585,25170. .25272,25803. .25951)
CDS

```

3775 GATGTGGG-TATTTCCGGCTCTCTGACATGTGAAGTGAAGCTGCTGTGCA 3717
 34553 TGTGAAAAATGTCGAAAAATGCA 34576
 3716 TGTG-TAAAAATGTCGAAAAATGCA 3694

RESULT 14
 LOCUS CELC04F5 39752 bp DNA INV 29-OCT-1998
 DEFINITION Caenorhabditis elegans cosmid C04F5.
 ACCESSION AF101305
 NID 93806127
 VERSION AF101305.1 GI:3806127
 KEYWORDS
 SOURCE
 ORGANISM
 . Caenorhabditis elegans.
 . Caenorhabditis elegans
 . Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 . Rhabditina; Rhabditoidea; Rhabditiidae; Peloderiinae; Caenorhabditis.
 1 (bases 1 to 39752)
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
 Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
 O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
 Saunders, D., Showkneen, R., Smaildon, N., Smith, A., Sonhammer, E.,
 Staeden, R., Sulston, J., Thierley-Mieg, J., Thomas, K., Vaudin, M.,
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
 Wilkinson-Spratt, J. and Wohlman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans

TITLE
 JOURNAL Nature 368 (6466), 32-38 (1994)
 MEDLINE 94150718
 REFERENCE 2 (bases 1 to 39752)
 AUTHORS Latreille, P., Bradshaw, H. and Elliott, G.
 TITLE The sequence of C. elegans cosmid C04F5.
 JOURNAL Unpublished (1998)
 REFERENCE 3 (bases 1 to 39752)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (26-OCT-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 e-mail: twenematode.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is Y58A7A, 200 bp overlap; 3' clone is F14P9, 2900 bp
 overlap. Actual start of this clone is at base position 1 of
 CELC04F5; actual end is at 38256 of CELC04F5

NOTES:

FEATURES

Coding sequences below are predicted from computer analysis, using
 the program GeneFINDER (P. Green and L. Hillier, ms in preparation).
 Location/Qualifiers

FEATURES	source
1..39752	/organism="Caenorhabditis elegans"
	/strain="Bristol N2"
	/db_xref="taxon:6239"
	/clone="C04F5"
	/chromosome="V"
	/complement(4336..6898)
gene	/gene="C04F5.2"
	/complement(join(4336..4484,5076..5296,5357..5472, 5544..5707,6181..6273,6391..6560,6612..6748,6815..6898))
CDS	/gene="C04F5.2"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="AAC69202.1"
	/db_xref="pid:93806129"
	/db_xref="gi:3806129"
	/translation="MYSVHATSVASSSLNCTSVMLRNHFIVAVNDKIAQISGLY ALLILVESLFFLLTTFVFRKRRSHPRFGILFGCVILCYIIGVAEVCQELFPG SSDLSVLYLIDLVAAFTLVFPCCLDIIRNKRKIDYIFCKLYCRKESQVEFLY YKRTIATWAVSYETRTWMLFLSVCLSVGTGYTFASFRCGIVSAVYITIGGC IALYFLNRKWTLYKQDISFPGKSLSMRYOIIENMISVRRLALPVITLSDLVDA LSKVTYTKIEQSNMCMYSMNYIPIYILNLHAWVEYSIPLYMLFRKAYSRLPNVC HRIPLRIQOSRESMIONVGEKELNVPTQAEYFAKLQNEWK"
	/complement(7985..9484)
gene	/gene="C04F5.3"
	/complement(join(7985..8175,8663..8727,8773..8887, 8946..9021,9072..9197,9267..9341,9449..9484))
CDS	/gene="C04F5.3"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="AAC69204.1"
	/db_xref="pid:93806131"
	/db_xref="gi:3806131"
	/translation="MLSQLPDLQAPGDTCSVTFQVKILNDQNNNSVPIDFHLPSGL IKNCVATRKRAVISSTIEFTDGRKTLKFEFRDEMFAKRLDELRLQNLVITEEK FAGSSAVFESDSSVIFSAPILOKYCEBRINVTLSVNFDFPIIMFSPENDVQYG PKSNFYICERSRKRTISESLQHRSTVFCGVVALSSIAHIVGHMLRRHMPHRELKYE NLDKLPAN"
	/complement(10719..12738)
gene	/gene="C04F5.4"
	/complement(join(10719..10912,10970..11171,11228..11335, 11385..11459,11875..12262,12680..12738))
CDS	/gene="C04F5.4"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="AAC69205.1"
	/db_xref="pid:93806132"
	/db_xref="gi:3806132"
	/translation="MPNEPDCQVMAEVAATSSFLRTSLAGMTLLCLICVCTIYSFLAI OKATKLFHNSKCIFFHTLIPAFIHMTIVRIILHGLDLINLVGPMSCGEIPSSRCEL RKLYKIAFVETPVPVITAEFRVAFRRARHNRKMGCVNLINPHISIALFELFIO SSEKVGSDIILYCMWSTGRKMLNPIPIYVSCQIITPALLYLILKRNKPEASIO KRSFLSGRYLSONLOTLKFRIVSTVYIYVYNAATFCVHFEFLAKSSAGQEPFI EIVACLPLYLLILFIVNWKEDKKPRHROFSIKVDYHPQYFNDLQKFFDOSFVKXKTK SVFPVVS"
	/complement(13663..15655)
gene	/gene="C04F5.5"
	/complement(join(13663..13856,14037..14238,14292..14399, 14454..14528,14601..14988,15594..15655))
CDS	/gene="C04F5.5"
	/note="contains similarity to C. elegans chemosensory receptor SRA-10 (GB:237092)"
	/codon_start=1
	/evidence=not_experimental
	/protein_id="AAC69206.1"
	/db_xref="pid:93806133"
	/db_xref="gi:3806133"
	/translation="MSLNEDCCQIMAEIATSSFLRTVLSSTTFLICICIPFCPLFAWL IHHRLHFNSKCIFFVTFNISVLVHVSRIILHGKDLINLVGPMQNCLEFPDRSRCD

```

/note="contains similarity to C. elegans chemosensory
receptor SRA-10 (GB:I237092)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69207.1"
/db_xref="PID:g3806134"
/db_xref="GI:3806134"
/translation="MPNNTDEOMATVATSSPELRAVLIAVLCLIMCIPISISLMLR
YFSVKLHFNKSLVLTNNTEFVLHCLARIVLKGKDLINFDWSSCDLIPSRARCH
RFLYRRGEVYIIEISPTILIERLAFKADYSENRKLIGLIEFTTHSTANRLSV
LPIFFLFTQLATIPGLLYLKRNEBELRASSLHKSTLTERYOISENLTSTMFIM
VWLVVVAATKAIGSYMMPYIESVLIPDLAITEVHICPLLYIIAALITVRDRPR
EFVNNYNEHHHEFELKFEFGDFAFERKOQVGPMSGKNKIVP"
complement(18816..21387)
/gene="C04F5.7"
complement(join(18816..19209,19266..19591,19641..19760
19807..19954,20000..20220,21076..21387))
/gene="C04F5.7"
/note="contains similarity to UDP-glucoronosyl and
UDP-glucosyl transferases (pfam: UDPGT_hmm, score: 57.39)
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69208.1"
/db_xref="PID:g3806135"
/db_xref="GI:3806135"
/translation="MYSHDFHMRIFAEATFLKHAERARALLTVMDGRSHATISIN
FHAIRKQDNHTTALFEFAHRODIDGMDERFIDMSGFONPESPQFDKIAFEDYSF
HOAVAYGESOTCNILKHKRREFEILEYEMDITLSLPAICGMAATSGKPRVY
NSTDLEAAOGSEKGFESRNVAITVPSNLPEFLMNTAYVYHHRMAADMGCSFTFT
YSTNFAOKMALESLIPRXPYSAEYENRRSSTFTDMPDLSLPPASRINDFSFGAYCK
ESKRRLDEPKFTIEHPKSKGTLIAFTFTDMRKAPKNYDARATVNRUSLEIRIM
SKGGRPGLKHKVATSSWVPQNDLIHNKTVLLSHGSLSTKEVISCAPITFVPPV
FGEORNAMLIKKEGFARINMKFKINDLDTDHMRVLEHNYOONAKFETVYMDQ
IPTLDEGAFKFNRLTKYXGKMPSHFPEKSLDSYFTVLNVDLLLTVPIMLYLWK"
complement(22210..24185)
/gene="C04F5.8"
complement(join(22210..22572,23161..23248,23629..23876,
24045..24185))
/gene="C04F5.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69201.1"
/db_xref="PID:g3806128"
/db_xref="GI:3806128"
/translation="MMDRGARKLDRIRIAYLRKADVPSKAPSKLRMPTEGFSDCTE
LNTSEDFLEDDVEELVERDELSTNGVCSRSRDTVPLNFISLSLOVYFLD
LMLPTCOHVLDIGSRIGAIYVGASIFITGNVSTIGVEMDSYDIALRLTIRGELKN
IEIRLDIRHLIAHVKFNQEIFVNNVNSFELSKDEQCEMFLHTHRRPCTVLAHNP
IEKTEHLDSFIEDWLEFVNTTECLEYEGANDELIEEMASICKIYIVDSRAHDS
VPE"
complement(25408..29777)
/gene="C04F5.9"
complement(join(25408..25796,26398..26911,28047..28538,
29777..30000))

```

Note: remainder of annotations omitted.

Query Match	10.5%	Score 105	DB 22	Length 39752
Best Local Similarity	78.4%	Pred. No. 1.8e-46		
Matches 207	Conservative 0	Mismatches 48	Indels 9	Gaps 6

```

Db 34313 ATGAGGTATTATCGGTGTCAGTAAATGCAAAATTCATTTTATATACATTGCT 34372
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3948 ATGGGGTTAATCAATGATGTATTGAAAATAACAAAATG--TATTTTAATACATT--TCC 3892
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34373 CCTATTACGTATGCGAAAAAAATGTTTGATATCTCATGTCATTTTGGCATTTTTC 34432
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3891 CCTATTATTTGGCGAGAAAG--TGTAATAAAA-ACGCATGTCATTTTATCATTTATTC 3836
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34433 GACATTTTTCATATCGCCGACATCGCATGTCATGATGGAAGGCGAGACGCAGAGA 34492
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3835 GACATTTTTCATATCGACCGAGATCCCATTTTCATCATGCTGAGAGACGACGACAAAA 3776
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34493 AATAAAACTCCCTCTGCGTCTCTCGGCTTCAGCATGTGAAGTGAATGTCTGCGTGA 34552

```


Query Match 14.1% Score 141: DB 19: Length 110000;
 Best Local Similarity 82.0% Pred. No. 2.62e-70;
 Matches 219: Conservative 0; Mismatches 42; Indels 6; Gaps 5;

Db 44711 GAAAGATATATGGGTTAATCATGATGACGAGAAATGCAAAA-TGATTTTCATACA 44769
 Cp 3957 GAAAAATATATGGGTTAATCATGATGATGAAATATCAAAAATGATTTTATACA 3898

Db 44770 TTTTCCCCCATTCATGATGATGAGAAATGTTTTTAAAAAAGCATGACATTTTCGACA 44829
 Cp 3897 TTTT-CCCCCATTCATGATGATGAGAAATGTTA-AAAAAGCATGATTTT-ACA 3842

Db 44830 TTTTTCATATTTTTCATGATGACGAGTCCACTTTCATGCTGAGAGACGAGACG 44889
 Cp 3841 TTTTTCATATTTTTCATGATGACGAGTCCACTTTCATGCTGAGAGACGAGACG 3782

Db 44890 CAGAGAAATACACACATCTCTGCGCTCTGCGCTTTCAGCATTTAAAGTGTGTCG 44949
 Cp 3781 CAGAAAGATGTGTGATTTTCGCGCTCTGCTGCTTTCAGCATGAGTGTGTCG 3722

Db 44950 GCCGATGTGAAAAATTTTGGAAAAATG 44976
 Cp 3721 TTTGATGTA-AAAAATGTGAAAAATG 3696

RESULT 11
 LOCUS CEY62E10 256941 bp DNA HTG 12-MAR-1999
 DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
 clone Y62E10, WORKING DRAFT SEQUENCE.
 ACCESSION AL031580
 NID 94469033
 VERSION AL031580.4 GI:4469033
 KEYWORDS HTG: HTGS PHASE1
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 256941)
 LLOYD, C.
 Direct Submission
 Submitted (12-MAR-1999) Nematode Sequencing Project, Sanger Centre,
 Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
 Washington University, St. Louis, MO 63110, USA. E-mail:
 jesus@anger.ac.uk or rwenematode.wustl.edu
 On Mar 22, 1999 this sequence version replaced gi:4455366.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n's separate
 segments.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1..256941
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="Y62E10"
 BASE COUNT 77397 a 44556 c 43039 g 79104 t 12845 others
 ORIGIN

Query Match 14.1% Score 141: DB 19: Length 256941;
 Best Local Similarity 82.0% Pred. No. 2.62e-70;
 Matches 219: Conservative 0; Mismatches 42; Indels 6; Gaps 5;

Db 112946 GAAAGATATATGGGTTAATCATGATGACGAGAAATGCAAAA-TGATTTTCATACA 113004
 Cp 3957 GAAAAATATATGGGTTAATCATGATGATGAAATATCAAAAATGATTTTATACA 3898

Db 113005 TTTTCCCCCATTCATGATGATGAGAAATGTTTTTAAAAAAGCATGACATTTTCGACA 113064
 Cp 3897 TTTT-CCCCCATTCATGATGATGAGAAATGTTA-AAAAAGCATGATTTT-ACA 3842

Db 113065 TTTTTCATATTTTTCATGATGACGAGTCCACTTTCATGCTGAGAGACGAGACG 113124
 Cp 3841 TTTTTCATATTTTTCATGATGACGAGTCCACTTTCATGCTGAGAGACGAGACG 3782

Db 113125 CAGAGAAATACACACATCTCTGCGCTCTGCGCTTTCAGCATTTAAAGTGTGTCG 113184
 Cp 3781 CAGAAAGATGTGTGATTTTCGCGCTCTGCTGCTTTCAGCATGAGTGTGTCG 3722

Db 113185 GCCGATGTGAAAAATTTTGGAAAAATG 113211
 Cp 3721 TTTGATGTA-AAAAATGTGAAAAATG 3696

RESULT 12
 LOCUS CELCED3A 7653 bp DNA INV 23-FEB-1994
 DEFINITION Caenorhabditis elegans cell death protein (ced-3) gene, complete
 cds.
 ACCESSION L29052
 NID 9456416
 VERSION L29052.1 GI:456416
 KEYWORDS cell death protein: interleukin-1 beta converting enzyme.
 SOURCE Caenorhabditis elegans (strain N2) DNA.
 ORGANISM Caenorhabditis elegans
 Eukaryota; Eukaryota; Metazoa; Nematoda;
 Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea;
 Rhabditidae; Caenorhabditis.
 1 (bases 1 to 7653)
 Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.
 The C. elegans cell death gene ced-3 encodes a protein similar to
 mammalian interleukin-1b-converting enzyme
 Cell 75, 641-652 (1993)
 94061982
 FEATURES
 source
 1..7653
 Location/Qualifiers
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 2167..2366
 /gene="ced-3"
 /number=1
 join(2232..2366,2430..2575,2854..3107,4303..4634,
 5547..5760,5815..5942,6298..6537,7013..7075)
 /gene="ced-3"
 join(2232..2366,2430..2575,2854..3107,4303..4634,
 5547..5760,5815..5942,6298..6537,7013..7075)
 /gene="ced-3"
 /note="codes for a protein similar to mammalian
 interleukin-1b-converting enzyme"
 /codon_start=1
 /product="cell death protein"
 /protein_id="AA027982.1"
 /db_xref="PDB:1A027982.1"
 /db_xref="GI:456417"
 /translation="MMRODRSLRLENNIMFSSHLKVEILEVLIAKOVLSNDGMI
 NSCGVREKREIRKAVORGDVAFDFADLRSTGHEGLAEVLEPLARSVDNAVE
 ECPMSRASHRSRALSPAGYSPTRVRHDSVSVSTSTODITSRAKRSRSLALS
 SDRAHYSPPVNAFSPSPSSANSSTFGSSLSIGSSSRNRFSKASGPTQYIFHEDMN
 FVDATTSRVFDEKTMFNESSPGKCLINNEHEDMPFNKTKADNDITNFRGM
 GYIVCKDNLTGKMLTIRDFAKHSHGDSAILVISHGEENVIGVDIPSTHET
 YDLNANAPRLANKPKIVFOACRGERBGDPVLDVSGVPAPFLRGMNRRGPEF
 NELGCVPOVOOVWRKKRPSODILIRVATTAQVSMRNSARGSWFIOAVCEVSTHAK
 DMDVLELLEYNKAVACGFOISQGSNLIKQPEWTSLSLKKTFWPARNSAV"
 2367..2429
 /gene="ced-3"
 /number=1
 /number=2
 /gene="ced-3"
 2430..2575
 /number=2
 2576..2853

intnon
 exon
 intnon

Matches	320:	Conservative	0:	Mismatches	52:	Indels	21:	Gaps	9:
Db	9519	CATTATTCGACATTTT	TACATCGACGACACATCCCATTCACATCCTGGAACGAGAGA	9578					
Qy	3696	CATTTCGACATTTT	TACATCGACGACAGCTCACTCACTGGAAGCGAGAG	3754					
Db	9579	CGCAGAAAGTG	TGTGTTCCTCCGCGCTGCTGCTTCACAGTCGTAAGGACCTGT	9638					
Qy	3755	CGCGGAGAAAT	TACACACATCTTCTTGCTCTCTCTGCTTCACAGCATGTGAAATGGGATCT	3814					
Db	9639	CGTTCGATGTAAAAA	-TGTCGAA-----AAATGCATGCGCTTTT	TACATTTTC	9688				
Qy	3815	CGGTCGATGTAAAAA	ATGTCGAAATGTAAAAAATGCATGCGCTTTT	TACACTTTC	3874				
Db	9689	TGCACACATAAAT	ATGAGG-AAAATATAT-AAAT-A-A-T-----GCATTTTTCGTCATCAT	9739					
Qy	3875	TGCACCAATGAAT	ATGAGGGAATATTAATTAATCATTTT	TGTATTTTTCACATCATC	3934				
Db	9740	ATGACACACACAT	GATTTTTCGTTGAGCAATTTCAAAAGTAGACATAC	TAGAACGAA	9799				
Qy	3935	ATGATTAACCCCAT	TATTTTTCGTTGAGCACTTAAAGTAGACATAT	TAGAGCGAA	3994				
Db	9800	AACCAAAATTTCT	ACAAGATATTCACATTTATGATATTTATGCTGTATTAATAGCATATC	9859					
Qy	3995	AACCAAAATTTCT	CAAGATATTAACCTTTATGATATTAATAGATTTAATAGCATATC	4054					
Db	9860	TTGAATGAAAAA	ATCAAAA-ATCTGCGAAAC	9891					
Qy	4055	TTGAATGAAAGT	CAGCAAAAATATGTCGAAAC	4087					
RESULT	8	CEV67H2	314495 bp	DNA	HTG	04-MAR-1999			
LOCUS		Caenorhabditis elegans DNA	** SEQUENCING IN PROGRESS **	from					
DEFINITION		clone Y67H2, WORKING DRAFT SEQUENCE.							
ACCESSION		AL022475							
NID		94469034	GI:4469034						
VERSION		AL022475.3							
KEYWORDS		HTG; HTGS_PHASE1.							
SOURCE		Caenorhabditis elegans.							
ORGANISM		Caenorhabditis elegans.							
REFERENCE		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.							
AUTHORS		1 (bases 1 to 314495)							
TITLE		McMurray, A.							
JOURNAL		Direct Submission							
COMMENT		Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1HQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@genes.ac.uk or rwenematode.wustl.edu							
		On Mar 22, 1999 this sequence version replaced gi:4468145.							
		IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.							
		* NOTE: This is a 'working draft' sequence.							
		* This record will be updated with the finished sequence							
		* as soon as it is available and the accession number will							
		* be preserved.							
FEATURES		Location/Qualifiers							
source		1..314495							
		/organism="Caenorhabditis elegans"							
		/db_xref="taxon:6239"							
		/chromosome="IV"							
		/clone="Y67H2"							
BASE COUNT		96542 a 54551 c 52760 g 97027 t 13615 others							
ORIGIN									
Query Match		14.2%; Score 142; DB 19; Length 314495;							
Best Local Similarity		81.4%; Pred. 5.60e-71;							

```

7          CEC48D1      39908 bp     DNA       INV        23-NOV-1998
LOCUS      Caenorhabditis elegans cosmid C48D1, complete sequence.
DEFINITION Z81049
ACCESSION  Z81049
VERSION    91627677
KEYWORDS   HTG.
SOURCE     Htg.
ORGANISM   Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittidae;
Rhabditoidea; Rhabdittidae; Peloderinae; Caenorhabditis.
REFERENCE  Burton,J.
TITLE      Direct Submission
JOURNAL    Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
AUTHORS    Jeesanger.ac.uk or tw@nemtoide.wustl.edu
           2 (bases 1 to 39908)
REFERENCE  Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
           Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
           Coulson,A., Craxton,M., Dear,S., Du,Z., Dudin,R., Favello,A.,
           Fulton,L., Gardner,A., Green,P., Hawkins,T., Hiller,L., Jier,M.,
           Jonsson,L., Jones,M., Kerhaw,J., Kirsten,J., Lister,N.,
           Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
           O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra.A.,
           Saunders,D., Showkeen.R., Smaldon.N., Smith,A., Sonhammer,E.,
           Staden,R., Sulston.J., Thierry-Mieg,J., Thomas.K., Vaudin.M.,
           Vaughan,K., Waterston,R., Watson,A., Weinstock.L.,
           Wilkinson-Sproat,J. and Wohlman.P.
TITLE      2.2 Mb Of contiguous nucleotide sequence from chromosome III of C.
           elegans
JOURNAL    Nature 368 (6466), 32-38 (1994)
MEDLINE    94150718
COMMENT    Coding sequences below are predicted from computer analysis, using
           predictions from GeneFinder (P. Green, U. Washington), and other
           available information.
           For a graphical representation of this sequence and its analysis
           see:
           http://webace.sanger.ac.uk/cgi-bin/display?db=wormacc&seq=sequence &obj=C48D1
Current Sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. This sequence is the entire insert of clone C48D1. The true right end of clone F5H82 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence Z81093. The end of this sequence (33020..39908) overlaps with the start of sequence Z82274.
location/Qualifiers
         1..39908
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="IV"
            /clone="C48D1"
            complement(94..4439)
               /gene="C48D1.1"
               complement(join(94..1020,1345..1951,2283..2580,2632..2765,2882..3147,3615..3739,4089..4431,4382..4439)))
                  /note="Similar to C.elegans zyg-11 like protein. cDNA EST EMBL:Z14909 comes from this gene. cDNA EST EMBL:T01768 comes from this gene"
                  /codon_start=-1
                  /protein_id="CACB02846.1"

```

```

/db_xref="PID:93924776"
/translat="MTEDGVMIKIDASTSTIKPEPADVILETHNEOVVYKSMIAF
GDPNPLTYNRDLGADWGIYDCSDSPGPGVYKAFIYVNDLVYQESGNLFVYGS
RLKFLANKPTSDPMKLEIFEDNRLEELSPKMSDPOVIAIILEDEENSAICYS
EOLSESSIKVHLPVSPAKAGOLITFEYVENDYTPDRPAIEYRELAINTIDSEAN
LISAKAACSITHTITYPQSAFADTFKGVVAAASMERGAYRELAINTIDSEAN
NVPLKISDGEKTEENDEAPLREKINDEPELOGTMEDEPELAEVEPVAAPAEPA
DPEVEAKIEENGVNDESLTOSDSVFPLOIESSASSETTESASAPKPKPKE
GRMPQSISESSSTISSKVDVSVIYVNDDECVGTEETLOKMTICTGAYVPL
RPGIFQFIAGNNYKTSIOTASARYSECEGFSDDVPAISANVITSSPKHSY
NKEVGSDRGVEIPIRTGDEFTGYSTQIMVOYGPDLACFAVAPAKROFLD
LDHLYOMKEDYNIOLSKRGKISLEPNEBNSKSGSTKRTFTPVGTGIVY
DMERKGVPCGVTCFQDVYKSTPFSOGGVPGVDLYNIEVELISGTDOCKMG
ASMTKEXPASLSTTVIRNNGEYKATILMKPELITYSEDNREYEHDOLGIVVDSI
YSDNEQSKVEFERELNPLITLFTTRILMKCYMAARLPIHREPSPPRIKYNEN
VYDFOKGGPOSVNSKSDROYOKEPMDQSYHDPADTISHSUNHMETRSNAG
TSVSYVDRKEVGVYMKVDKNTQIVLSGKAAAHMMHNTSOSTIVGNSYIFMNP
EVHKAHYAPEITQVVEIINTENHGVDSSESLFEDLCWDRYVERLAFEAL
VLRNDGVVMASSNVTPLNPSMTAVRENTDRNLRGENTIEEYREDISIOKRIAH
DGHSDOPVSAAYEELHEITATSNAYIVRENTDKRKRESENLVGYCRYLKYKPREIV
MNEGONQOQOOSLSRGPTRPSYSPGSSQSHONDNRNSAHNGDSSTQSHSA
GGSNNHONFYQPVYQSSSQOSTQPMINBAPSGPDPSPSSQSTFNGCAPPPYXSMS
NDYNRMYRLAEITDAOMIINNDIQIKNISPLDLDERIDKQREKHYEK
YGN"
complement(27575..32349)
/gene="F5262.3"
complement(join(27575..27979,29638..29814,29857..29959,
30010..30132,30177..30380,30607..30741,31300..31446,
31498..31803,32240..32349))
/gene="F5262.3"
/codon_start=1
/protein_id="CAB05205.1"
/db_xref="PID:el354462"
/db_xref="PID:g3924775"
/db_xref="GI:3924775"
/translat="METRLVLYKDIILNREAIICLNRETHELLNANVESVOPQIGQL
AVENLKIVQDVCFISPECCTIIEPAPDCVILPEKEDVMEFKSMGFSDDSH
LYDENIGYADWGVNCSQKAKNTKVAEPKYLAVNDLCKVFLKMAVQLVY
IDNRPAEVLKMEFERRIADIORTKORLEIYAVPIIPSAHPSNHSYTGPE
SCDSLRPOPVPAARLTSROPTNOAPREIOORSYNSIIEGSGRTPOPE
TANSKRGTEYISGRSSSDOYIRNEPPLPAPILPSPMGMYENGMIOYLRVY
OKAIAHYSKQIITISLVFNCSKPIPVQATIHMAFETAINNNVNNENKRYTG
KECFALFYVDSLREPKFSKMTTIGNPIFSCELVGYMALHSJGTDJNKRNDI
EMOKKRAOQEHFAOQRPAGSSNYOTENAOHSNNMYRHSQAPVPSRSSISLG
ETAEFKNLERTNDMESLCAIKKFLDNQLKETLRNGYRNRQLALEKLECSHLL
EKN"
complement(33889..34904)
/gene="F5262.4"
complement(join(33889..33909,33958..34013,34063..34141,
34714..34802,34853..34904))
/gene="F5262.4"
/notes="CDNA EST YK356f5.5 comes from this gene; CDNA EST
EMBL:U64203 comes from this gene; CDNA EST EMBL:U66267
comes from this gene; CDNA EST YK298f3.3 comes from this
gene; CDNA EST YK356f5.3 comes from this gene; CDNA EST
YK383e11.3 comes from this gene"
/codon_start=1
/protein_id="CAB05208.1"
/db_xref="PID:el354465"
/db_xref="PID:g3924777"
/db_xref="GI:3924777"
/translat="MRGVEBNDGLVMTMCRVVEKKDKENKMRDESGRIIDKHNH
TMGELDTLLRDPNCGRHMSRVEKMPGTDRFDVNRQYKEVYESIYANY"
BASE COUNT 12956 a 6666 c 6504 g 13145 t
ORIGIN
Query Match 15.4%; Score 154; DB 21; Length 39271;
Best Local Similarity 85.7%; Pred. No. 4.67e-79;
Matches 221; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
Db 3227 CATTTTCTACATTTTCTACATGACGACGACATCTCTCATGCTGAAGACGAGAGA 3286

```

```

OY 3696 CATTTTGGACATTTT-TACATGAAAGACAGCTCACTTACATGCTGAAGACGAGAGA 3754
DB 3287 CGCAGAAATACCCACACATCTCTCGCTCTCGCTTCCAGCATTTAAAGTAGCTGT 3346
OY 3755 CGCGAGAAATACCCACACATCTTCTCGCTCTCGCTTCCAGCATGGAATGGATCT 3814
DB 3347 CGCGCATGTAGAAAATTTGAAAAATGTGAAAAATGATGCTGTTTTTAAAAACATT 3406
OY 3815 CGGTCAGTGTAAAAAATGTCGAAATTAATGTA-AAAAATGATGCGTTTTTTTACA-C- 3870
DB 3407 TTTTACACACATGAAATTAAGAAAAGTGTAAATATACATTTT-CTATTTCGTCA 3465
OY 3871 TTTTGCACAAATGTAATGAGGGAATATGATTAATATACATTTTCTATTTTTCAACA 3920
DB 3466 TCACATGAATAACCCCAT 3483
OY 3931 TCACATGATTAAACCCCAT 3948

RESULT 5
LOCUS CEY45F10 434238 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** From
clone Y45F10, WORKING DRAFT SEQUENCE.
ACCESSION 293245
NID 93378089
VERSION 293245.1 GI:3378089
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 434238)
AUTHORS McMurtry,A
TITLES Direct Submision
JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RO, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@anger.ac.uk or twenematode.wustl.edu
COMMENT
On Aug 4, 1998 this sequence version replaced gi:1906319.
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
location/Qualifiers
source 1..434238
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y45F10"
BASE COUNT 141590 a 73465 c 71957 g 143994 t 3232 others
ORIGIN
Query Match 15.4%; Score 154; DB 19; Length 110000;
Best Local Similarity 85.7%; Pred. No. 4.67e-79;
Matches 221; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
Db 44828 CATTTTCTACATTTTCTACATGACGACGATCTCATGCTGAAGACGAGAGA 44887
OY 3696 CATTTTGGACATTTT-TACATGAAAGACAGCTCACTTACATGCTGAAGACGAGAGA 3754
DB 44888 CGCAGAAATACCCACACATCTCTCGCTCTCGCTTCCAGCATTTAAAGTAGCTGT 44947
OY 3755 CGCGAGAAATACCCACACATCTTCTCGCTCTCGCTTCCAGCATGGAATGGATCT 3814
DB 44948 CGCGCATGTAGAAAATTTGAAAAATGTAAGAAAATGATGCTGTTTTTAAAAACATT 45007

```



```

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="1Y"
/clone="F5262"
/feature="join(2363, .4916)
/gene="F5262.1b"
/complement(2363, .5967)
/gene="F5262.1a"
/complement(join(2363, .2713,2762, .2938,4410, .4841,
5941, .5967))
/gene="F5262.1a"
/note="Similar to Bacterial mutr protein: cDNA EST
yk332h11.5 comes from this gene: cDNA EST yk4281h9.5 comes from this
gene"
/codon_start=1
/protein_id="CAB05206.1"
/db_xref="PID:el354463"
/db_xref="PID:g3947559"
/db_xref="GI:3947559"
/db_xref="SP:REML.062257"
/translation="MASNSTNSKKKSTEEPPSVQKLKLLAQAKNSDSEQST
SKRPKRRKRAAQAQASAPAGEPEKKKOPKRAVSGMQLQAKNSISOTRRQ
VSTSGSSNRKTPAEQOQYKGRPIPIIDLEFRIISNVCETINDNIR
CFHEILAHMYIDHVEMDKISCPNWSGSDENQCHRYIKRAHADEYLAKR
EYESTVAYTALVDEMDHVLVQSYFAKKGWGPCKKINDAEPDRDAIREFEE
TGDFGCIYSKEKKRPFINDGNVRLILVKNVPKDNFQPTREIRFEFLD"
/complement(join(2363, .2713,2762, .2938,4410, .4916))
/gene="F5262.1b"
/note="Similar to Bacterial mutr protein: cDNA EST
yk357h8.5 comes from this gene"
/codon_start=1
/protein_id="CAB05204.1"
/db_xref="PID:el354461"
/db_xref="PID:g3924774"
/db_xref="GI:3924774"
/db_xref="SP:REML.062255"
/translation="MSTENSKKPKRNSIFSKNISQKNSKEEPPSSVQKLK
LQAKNSDSEQSTSKRPKRRKRAAQAQASAPAGEPEKKKOPKRAVSGRMQ
LOAKNRISOTRRQVSTSGSSNRKTPAEQOQYKGRPIPIIDLEFRIISNVCETINDNIR
CFHEILAHMYIDHVEMDKISCPNWSGSDENQCHRYIKRAHADEYLAKR
EYESTVAYTALVDEMDHVLVQSYFAKKGWGPCKKINDAEPDRDAIREFEE
TGDFGCIYSKEKKRPFINDGNVRLILVKNVPKDNFQPTREIRFEFLD"
/complement(join(2363, .2713,2762, .2938,4410, .4916))
/gene="F5262.1b"
/note="Similar to Bacterial mutr protein: cDNA EST
yk357h8.5 comes from this gene"
/codon_start=1
/protein_id="CAB05207.1"

```


REFERENCE
AUTHORS
2 (bases 1 to 39908)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, D., Percy, C., Rikken, L., Roopra, A., Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaubin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkerson-Sproat, J. and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*
Nature 368 (6466), 32-38 (1994)
94150718
JOURNAL
MEDLINE
COMMENT
Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:
http://webc.sanger.ac.uk/cgi-bin/display?db=wormaceclass=Sequence&object=C48D1
Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
This sequence is the entire insert of clone C48D1. The true right end of clone F58D2 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence Z81093. The end of this sequence (33020..39908) overlaps with the start of sequence Z82274.
FEATURES
Source
Location/Qualifiers
1..39908
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(94..4439)
/gene="C48D1.1"
complement(join(94..1020,1345..1951,2283..2580,2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/gene="C48D1.1"
/note="Similar to *C. elegans* ZYG-11 like protein; CDNA EST EMBL:Z14909 comes from this gene; CDNA EST EMBL:101768 comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="PID:e1344741"
/db_xref="PID:g3875048"
/db_xref="GI:3875048"
/db_xref="SPRMBL:002227"
/translation="MRKPLHDIACOKIAEFVONGSYNREFELDSKNNTVYILNRDKNPLKTIHKLIVNTKYTLHPDILDEMTAQLKOPISKEIKTGFSSKSWCIEFRKE TRHYDIVRILEAVNETRKNLTOLHLEKOKLMKRWTAIGMPLSHLTHLSDVH NEDDFYOLFGFNSLRSLDLSGNIINIAISRLKOLEVALDFSAITLDELKDFN LKHLRVLDVSQTRFNLSNNNIVYEOFECKMEPELPEFMCTSTSGSLDILKLTFO LSPINISVTEAETSYNDVRLNATFAEFHESYVALYCLAREPSPGCMKARISL SEDNLKLNIPSCIKTIYKSMOTFDSISYVONKLCILFGMIXSQDLSNGTYLIVE ALMSFVLYKRMISEAVYKYNISYMAVEKLVMLKGLDFDKISTAMSFLEENFVS KALTRASSVLOCICIRHMSDELKSEKTEENNRVLDEMSVGOKNFKSEFEPI TIKRTKSSKACACTNMIOAGIISLSICMEKVESNDITTKLELITMISALNVDISQE LYTSSNDYKIGQANAYSTEK"
complement(6258..11120)
/gene="C48D1.2"
complement(join(6258..6349,6876..7065,7420..7547,7602..7815,8728..9059,10248..10501,10780..10925,10989..11120))
gene
CDS

/gene="C48D1.2"
/note="Similar to Caspase recruitment domain, ICE-1 like protease (caspase) p20 domain, ICE-like protease (caspase) p10 domain; CDNA EST EMBL:D75533 comes from this gene; CDNA EST yk198f10.5 comes from this gene; CDNA EST yk491a5.5 comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="PID:e1344743"
/db_xref="PID:g3875050"
/db_xref="GI:3875050"
/db_xref="SPRMBL:002229"
/translation="MRDRRSILRNIMFSSHLKXDELLEVIYIAQVUNSDGMIN SCGVREKREIRVAVORRGDVAFDAYDLARSTGBGLAEVLEPLARSDNAVEE CRNPASHRRSRALSPAGTISPTVRHDSVSVSTSTYODIYSRARSRSRSLALS DRHYSPPVNAFSPSPSSANSSTFGCSSLGSSSSNRSPSKACPTQYIFHEDNF VDAPTIISRVDEKTYMNFNSPFGMCILINHEFEDMPFNGTKADKNDLNFRCMG YTVACKNDLGRGMLITRDFPAKHESHGDAIIVLISHGEENVITGVDIPISITHTY DLNAANAPRLANKPKPIFYVACRGERRDNGFPVLDVSGVAFELRGKDNDRDPLEN FLGCVRQNOQVWRKRPQSDIITATTAITTAIVSMKNSAGSWFIOAVCEVSTHAKD MDVVELTEVKNKVAIDIPPAQVLLAGSTKLRUKFTRDSLPN"
complement(16940..21616)
/gene="C48D1.3"
complement(join(16940..17126,18419..18597,18985..19442,19553..19879,19925..19961,20333..20481,20789..20987,21037..21152,21373..21616))
/gene="C48D1.3"
/note="Similarity to *Salmonella* sodium/proline symporter (SW:PUTP_SALTY); CDNA EST EMBL:C13840 comes from this gene; CDNA EST EMBL:C11562 comes from this gene"
/codon_start=1
/protein_id="CAB02847.1"
/db_xref="PID:e1344742"
/db_xref="PID:g3875049"
/db_xref="GI:3875049"
/db_xref="SPRMBL:002228"
/translation="MADLIGIAIVPEVYLLIVGIMAGRKSKELESEGAATEE VMAGNEITLIGITPTMTWTWGAIVINSTALNGLGCGQAVGVAISLVMGGL FAKMRBEETITMLDPEQFQWNLLEIFGTFENFKRLGFLVLIILIDFEOHKG YORGLMAYPALGFTFWTAALISALGATSVILIDIMNASVTLASCIAYFTFGY YAVAYTDVOLFICFVGLIILGLYONRNRKERSLMDKRLTLVFGIPMOVFEOR VLSKTAHGAOTISFVAGVCIILMAIPALICAIRINDMRTDVSPPANNKTESIP PBRNNVAVLFOVLYLTPRNFATIGGANSAAVSSADSVSASAEKHWIKLIRP HASEKEVITVMAIITICVIMATITMALITIOSIYGLMYLCADLVVILPEOLLGVMP KSTTGLSLAGVAVGLRIGGEPLVSPAFEHYPMYDGVYFFPTFATMLSSMATI YVISOEKLEKFSRSPEDVMGCVNIPIDHVLDPDVSFAVSEFTLNKKECDGM QPPOLOTEHRLQYIRITNSRLMKIHYIHIRFVILIPOLAIRDELVPRVHLISQIFF CIFFSR"
BASE COUNT 13182 a 6945 c 7146 g 12635 t
ORIGIN
Query Match 95.6%; Score 956; DB 21; Length 39908;
Best Local Similarity 99.3%; Pred. No. 0.00e+00;
Matches 993; Conservative 0; Mismatches 1; Indels 6; Gaps 6;
Db 8762 TACCCATGATGTGAACAATTTGGTAAGATTGCTTGTGGCCTTGATACATTCGT 8821
|||
Db 4600 TACCCATGATGTGAACAATTTGGTAAGATTGCTTGTGGCCTTGATACATTCGT 4541
|||
Db 8822 GTTGGATCTGGCAAGGTTCAATTTATGATGAGGACACATTCACAGAGACTCGAG 8881
|||
Db 4540 GTTGGATCTGGCAAGGTTCAATTTATGATGAGGACACATTCACAGAGACTCGAG 4481
|||
Db 8882 AAGTTCTGTACATGTTCTTCGTGCAAAAACAGCGCTTATGTTGGTCAATCGCAAG 8941
|||
Db 4480 AAGTTCTGTACATGTTCTTCGTGCAAAAACAGCGCTTATGTTGGTCAATCGCAAG 4421
|||
Db 8942 TTCAATCTCTTCAATGCAATATGTAATGAGTTGGTCCAGAACTTTCTGAATAGCGA 9001
|||
Db 4420 TTCAATCTCTTCAATGCAATATGTAATGAGTTGGTCCAGAACTTTCTGAATAGCGA 4361
|||
Db 9002 TTACGACTGCACTATACCGGAGAAAGCATCGCGTGAATGAAGACTTGGCGATGCT 9061
|||

			/protein_id="AAA27982.1"
		/db_xref="PID:g456417"	
		/db_xref="GI:456417"	
		/translation="MNRQDRSLLEERNIMFSHKLKVEILLVLIKAVLYNSDNGMT NSCGVIREKREIIVKAVQRRGDVAFADALSTGHELAIVLEPLKARVSDNAVEE ECMPNSHRSRSLSPAGTSPTRVHRDVSVSSEFTYODIYSRARSRSRSLHJ SDRHNSYPVNPAPPSOPSSANSSEFTGSSLSGSSRNRSFSKASGPQYLFHEEDM FYDAPTISRVEDEKTYRNFSNPRGMCLIIINNEHPEOMPTRNGTKADNDLNLNRECK FYDIYCKNLTGRGMLTIIRDPAKHESHGSLALVILSGEENVLIIGVDLIPISHE YDLVLAANAPRLANKPKTIVEVQCRGRNGSPVLDSDGVPAELRGRMNDRODPLE NFLCGRPQVOOVWRKRPQADILIRATTAQVSWRNARSRSWFLQAVCEVFSHAH DMVDVELLTLEVNKKVACGFOTSQSSNLTLMKPMPTSLIKKFYFWPEARNASV"	
		2367..2429	
intron		/gene="ced-3"	
		/number=1	
		2430..2575	
exon		/gene="ced-3"	
		/number=2	
		2576..2853	
intron		/gene="ced-3"	
		/number=2	
		2854..3107	
exon		/gene="ced-3"	
		/number=3	
		3108..4302	
intron		/gene="ced-3"	
		/number=3	
		4303..4634	
exon		/gene="ced-3"	
		/number=4	
		4635..5546	
intron		/gene="ced-3"	
		/number=4	
		5547..5760	
exon		/gene="ced-3"	
		/number=5	
		5761..5814	
intron		/gene="ced-3"	
		/number=5	
		5815..5942	
exon		/gene="ced-3"	
		/number=6	
		5943..6297	
intron		/gene="ced-3"	
		/number=6	
		6298..6537	
exon		/gene="ced-3"	
		/number=7	
		6538..7012	
intron		/gene="ced-3"	
		/number=7	
		7013..7652	
exon		/gene="ced-3"	
		/number=8	
		1455 c	1271 g 2498 t
BASE COUNT	2429 a	1455 c	1271 g 2498 t
ORIGIN			
Query Match	100.0%;	Score 1000;	DB 21; Length 7653;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;	
Matches 1000; Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
Db	3601	TCATGTGTTGTCATATGATGAGAGAAATAGCAATTTTATATATTTCCCAT	3660
Oy	3601	TCATGTGTTTGTTCATATCAATGACGAAATATAGCAATTTTATATATTTCCCAT	3660
Db	3661	TCATGTTGTGCGAAAAATAGTAAAAAGGCGATGCAATTTTGCACATTTTTCATCGA	3720
Oy	3661	TCATGTTGTGCGAAAAATAGTAAAAAGGCGATGCAATTTTGCACATTTTTCATCGA	3720
Db	3721	ACGACAGCTCACTTCATACATGCTGAAGACGAGAGACCGGAGAAATACACACATCTTCT	3780
Oy	3721	ACGACAGCTCACTTCATACATGCTGAAGACGAGAGACCGGAGAAATACACACATCTTCT	3780
Db	3721	ACGACAGCTCACTTCATACATGCTGAAGACGAGAGACCGGAGAAATACACACATCTTCT	3780
Oy	3721	ACGACAGCTCACTTCATACATGCTGAAGACGAGAGACCGGAGAAATACACACATCTTCT	3780

Db	3781	GGGTCCTCTGCTTTCACAGCATGTGAATGGGATCGCGGTGATGTGAATAAATGTCGATA	3840
Qy	3781	GGGTCCTCTGCTTTCACAGCATGTGAATGGGATCGCGGTGATGTGAATAAATGTCGATA	3840
Db	3841	ATGTAAAAAATGCATGGGTTTTTTTACACTTTCTGCAAAATGAATAGGGGAAAAATGT	3900
Qy	3841	ATGTAAAAAATGCATGGGTTTTTTTACACTTTCTGCAAAATGAATAGGGGAAAAATGT	3900
Db	3901	ATTAAATACATTTTTTGTATTTTTTCAACATCATATGATTATACCCATTAATTTTTGGTT	3960
Qy	3901	ATTAAATACATTTTTTGTATTTTTTCAACATCATATGATTATACCCATTAATTTTTGGTT	3960
Db	3961	GAGCACTTAAAAAGTAGAGAAATATTAGAGCGAAACCAAAATTTCTTCACAGATATTACC	4020
Qy	3961	GAGCACTTAAAAAGTAGAGAAATATTAGAGCGAAACCAAAATTTCTTCACAGATATTACC	4020
Db	4021	TTTATTGATTAATTTAGATGTTTAATAGCATATCTTGAATTAAGTCAAGCAAAATATGT	4080
Qy	4021	TTTATTGATTAATTTAGATGTTTAATAGCATATCTTGAATTAAGTCAAGCAAAATATGT	4080
Db	4081	GGGAACACCTGAAAAAATCAAAAATTCCTCGAAAAATTTGAAAAATGCAATTAATACA	4140
Qy	4081	GGGAACACCTGAAAAAATCAAAAATTCCTCGAAAAATTTGAAAAATGCAATTAATACA	4140
Db	4141	TTTTGCATTTTTTCTACATCATATGATGTAGAAAATTTAAAAAGGAAATCAAAATTTCTA	4200
Qy	4141	TTTTGCATTTTTTCTACATCATATGATGTAGAAAATTTAAAAAGGAAATCAAAATTTCTA	4200
Db	4201	GAGGATATATTGAATGAATGAACATTGCCGAATTTAAATGTGGGAAAAACGTCAAAAAAGAGA	4260
Qy	4201	GAGGATATATTGAATGAATGAACATTGCCGAATTTAAATGTGGGAAAAACGTCAAAAAAGAGA	4260
Db	4261	AATTGGGATCAAAATGATCTCTAAAACCAACACATTTTCAGATCCGCAACTCTTCAT	4320
Qy	4261	AATTGGGATCAAAATGATCTCTAAAACCAACACATTTTCAGATCCGCAACTCTTCAT	4320
Db	4321	TCACCGGATGCTCTTCTCTCGGATACAGTCAAGTCGTAATTCGTCATTGCGCAAGGCTT	4380
Qy	4321	TCACCGGATGCTCTTCTCTCGGATACAGTCAAGTCGTAATTCGTCATTGCGCAAGGCTT	4380
Db	4381	CTGGACCAACTCAATACATATTCCATGGAAGAGATTAACATTTGTGATGCATGCCAACCA	4440
Qy	4381	CTGGACCAACTCAATACATATTCCATGGAAGAGATTAACATTTGTGATGCATGCCAACCA	4440
Db	4441	TAAAGCGTGTTCGACGAGAAACCACTGATACAGAACTTCTCGAGTCTCTGTGGAATGT	4500
Qy	4441	TAAAGCGTGTTCGACGAGAAACCACTGATACAGAACTTCTCGAGTCTCTGTGGAATGT	4500
Db	4501	GGCTATCATTAATATATGACACTTTGAGACATGSCCAACAGGATGTTACCAAGGCGG	4560
Qy	4501	GGCTATCATTAATATATGACACTTTGAGACATGSCCAACAGGATGTTACCAAGGCGG	4560
Db	4561	ACAAGACAATCTTACCAATTTGTTCAGATGATGGGCTA	4600
Qy	4561	ACAAGACAATCTTACCAATTTGTTCAGATGATGGGCTA	4600
RESULT	2		
LOCUS	CEC48D1	39908 bp	DNA
DEFINITION	Caenorhabditis elegans cosmid C48D1, complete sequence.		
ACCESSION	N81049		
NID	91627677		
VERSION	281049.1	GI:1627677	
KEYWORDS	HTG.		
SOURCE	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis		
AUTHORS	Burton, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jesse@anger.ac.uk or rtw@emulode.wustl.edu		

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.".
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA PAULEY A., GATUNG S.,
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.,
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U53181; G1245690;
 SQ SEQUENCE 134 AA; 14819 MW; 4F1CB795 CRC32;

Query Match 1.4%; Score 7; DB 5; Length 134;
 Best Local Similarity 100.0%; Pred. No. 3.55e+00;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 35 YSSPPVN 41
 |||||
 QY 166 YSSPPVN 172

RESULT 15
 ID 065580 PRELIMINARY; PRT; 162 AA.
 AC Q65580;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE UL6 PROTEIN (FRAGMENT).
 GN UL6.
 OS BOVINE HERPESVIRUS TYPE 1.
 OC VIRUSES; DSDNA VIRUSES; NO RNA STAGE; HERPESVIRIDAE;
 OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCHOENBOEKEN;
 RX MEDLINE; 96135223.
 RA SCHMITT J., KEIL G.M.,
 RT "Identification and characterization of the bovine herpesvirus 1 UL7
 RT gene and gene product which are not essential for virus replication
 RT in cell culture.".
 RL J. VIROL. 70:1091-1099(1996).
 DR EMBL; X91751; G1006629; -.
 TI NON_TER 1
 SQ SEQUENCE 162 AA; 18656 MW; 4BA35B45 CRC32;

Query Match 1.4%; Score 7; DB 14; Length 162;
 Best Local Similarity 100.0%; Pred. No. 3.55e+00;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 100 RSRRSR 106
 |||||
 QY 150 RSRRSR 156

Search completed: Tue Aug 10 11:41:55 1999
 Job time : 109 secs.

```

AC 018749;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE PROTAGINE PL.
OS PLANTAGINAE SP. 1.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYROMORPHA; DASYURIDAE; PLANIGALE.
RN [1]
RP SEQUENCE FROM N.A.
RA KRAJEWSKI C., BLACKET M., BUCKLEY L., WESTERMAN M.;
RL MOL. PHYLOGENET. EVOL. 0:0-0(1997).
DR EMBL: AF001595; G2330935; -.
SQ SEQUENCE 62 AA; 8290 MW; 12E28334 CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 62;
Pred. No. 3.55e+00; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Db 9 RSRSR 15
Qy 150 RSRSR 156

RESULT 11
ID 011288 PRELIMINARY; PRT; 128 AA.
AC 011288;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 13.2 KD PROTEIN (FRAGMENT).
X1-55.
GN
OS MOLUSCUM CONTAGIOSUM VIRUS SUBTYPE 1 (MCV1).
OC VIRUSES; DSNNA VIRUSES; NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC MOLUSCIPPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA MORATILLA M., AGROMAYOR M., NUNEZ A., FUNES J.M., VARAS A.J.,
RA LOPEZ-ESTEBAN J.L., ESTEBAN M., MARTIN-GALLARDO A.;
RL VIRUS GENES 0:0-0(0).
DR EMBL: U86881; G2105171; -.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1 128
FT NON_TER 1 128
SQ SEQUENCE 128 AA; 13152 MW; 93E22F10 CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 14; Length 128;
Pred. No. 3.55e+00; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Db 4 RSRSR 10
Qy 150 RSRSR 156

RESULT 12
ID 019225 PRELIMINARY; PRT; 130 AA.
AC 019225;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F08H9.2 PROTEIN.
X1-
GN F08H9.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WILD A.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RX SEQUENCE FROM N.A.
MEDLINE: 94150718.

```

```

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STRADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
DR EMBL: 277657; E1345117; -.
SQ SEQUENCE 130 AA; 14710 MW; EFB5A123 CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 5; Length 130;
Pred. No. 3.55e+00; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Db 32 RSRSR 38
Qy 150 RSRSR 156

RESULT 13
ID 044177 PRELIMINARY; PRT; 134 AA.
AC 044177;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE ORE134.
OS SYNECHOCOCCLUS PCC7002 PR-6.
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCCLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC 7002; ATCC 27264;
RA AKIYAMA H., KANAI S., HIRANO M., SUGIMOTO M., KIYOHARA M.;
RT "Cloning and characterization of Rubisco large subunit and small
RT subunit from Synechococcus sp. PCC7002."
RL SUBMITTED (DEC-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: D13971; D1003583; -.
SQ SEQUENCE 134 AA; 15269 MW; 6C6F48FF CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 134;
Pred. No. 3.55e+00; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Db 102 RSLRL 108
Qy 6 RSLRL 12

RESULT 14
ID 020104 PRELIMINARY; PRT; 134 AA.
AC 020104;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE THIS GENE MAY BEGIN IN THE NEXT COSMID.
X1-
GN F36D4.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

```

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL NATURE 368:32-38(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA LANGSTON Z., WOHLIDMAN P., GILLAM B.,
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.,
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U97405; G2039375; -
 SQ SEQUENCE 327 AA; 36425 MW; 81BF02D6 CRC32;

Query Match 1.6%; Score 8; DB 5; Length 327;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 87 ARSRSR 94
 149 ARSRSR 156

RESULT 7
 ID 042840 PRELIMINARY; PRT; 456 AA.
 AC 042840;
 DT 01-JUN-1998 (TREMREL. 06, CREATED)
 DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 46.7 KD PROTEIN (FRAGMENT).
 GN SPAC23A1.01C.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIZOOCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA MURPHY L., HARRIS D.,
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA WOOD V., BARRELL B.G., RAJANDREAM M.A.,
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL021813; E1250574; -
 KW HYPOTHETICAL PROTEIN.
 FT NON_TER 456 456
 PT SEQUENCE 456 AA; 46687 MW; 9216E7DF CRC32;
 SQ

Query Match 1.6%; Score 8; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 307 SSVSFTS 314
 134 SSVSFTS 141

QY 134 SSVSFTS 141

RESULT 8
 ID 018207 PRELIMINARY; PRT; 619 AA.
 AC 018207;
 DT 01-JAN-1998 (TREMREL. 05, CREATED)
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMREL. 09, LAST ANNOTATION UPDATE)

DE Y48B1C.3 PROTEIN.
 GN Y48B1C.3.
 OS CAENORABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITIA; RHABDITIDEA; RHABDITIDAE; PELODERINAE; CAENORABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MCMURRAY A.,
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
 RA CRAYTON A., DEAR S., DU Z., DUREIN R., FAVELL A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL NATURE 368:32-38(1994).
 DR EMBL: 293394; E1354641; -
 SQ SEQUENCE 619 AA; 72914 MW; 5AA47D9C CRC32;

Query Match 1.6%; Score 8; DB 5; Length 619;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 195 HLKVDL 202
 20 HLKVDL 27

RESULT 9
 ID 077435 PRELIMINARY; PRT; 1045 AA.
 AC 077435;
 DT 01-NOV-1998 (TREMREL. 08, CREATED)
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE EG:34F3.4 PROTEIN.
 GN EG:34F3.4.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CATHERINE SALLES, PHILIPPE VALENTI, ARETI DARLAMITSOU,
 RA NAOLINE HENDERSON, LORNA CAMPBELL, DAVID GLOYER,
 RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA BENOS P.,
 RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL031583; E1321009; -
 SQ SEQUENCE 1045 AA; 116521 MW; 0EAD4D79 CRC32;

Query Match 1.6%; Score 8; DB 5; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 129 TRAKMDV 136
 447 TRAKMDV 454

QY 447 TRAKMDV 454

RESULT 10
 ID 018749 PRELIMINARY; PRT; 62 AA.
 AC 018749;

OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA LYNE M., RAJANDREAM M.A., BARRELL B.G., WHITHEAD S., CHILLINGWORTH T.,
 RA CHUGCHER C.M.;
 RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL031158; E1313583;
 FT NON_TER 174 174
 SQ SEQUENCE 174 AA; 20505 MW; 92ABDE3 CRC32;
 Query Match 1.6%; Score 8; DB 3; Length 174;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 145 VSARSRS 152
 146 VSARSRS 153
 QY

RESULT 4
 ID 013854 PRELIMINARY; PRT; 263 AA.
 AC C13854;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE SPAC19G12.17C PROTEIN (FRAGMENT).
 GN SPAC19G12.17C.
 OS SCHIZOSACCHAROMYCETES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA OLIVER K., HARRIS D.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: 297209; E1250596;
 KM HYPOTHETICAL PROTEIN.
 FT NON_TER 1 1
 SQ SEQUENCE 263 AA; 26445 MW; 846167CF CRC32;
 Query Match 1.6%; Score 8; DB 3; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 SSVSFTS 15
 134 SSVSFTS 141
 QY

RESULT 5
 ID 007016 PRELIMINARY; PRT; 301 AA.
 AC 007016;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 33.7 KD PROTEIN.
 GN YFER.
 OS BACILLUS SUBTILIS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SUBMITTED F.C.;
 RA SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERBERO M.G., BESSIERES P., BOLOIN A., BOSCHERT S.,
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRICH S.D., EMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALIERON N.,
 RA GHIN S.Y., GLASER P., GOFPEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUSEPPE G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HUILO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAMA K., OGIMARA A., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTELELLA D., PORTELLI S., PRESCOTT A.M.,
 RA PRESECAN E., PUTIC P., PURNELLE B., RAPPORT G., REY M., REYNOLDS S.,
 RA RIGER M., RIVOLTA C., ROCCH E., ROCHE B., ROSE M., SADATE Y.,
 RA SATO T., SCANTAN E., SCHLEICH S., SCHROETER R., SCOPFONE F.,
 RA SERIGUCHI J., SEKONSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDO B.,
 RA SEROKIN A., TACCONE E., TAKAGI T., TAKAHASHI H., TAKEHARA K.,
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASAROTTI A.,
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL NATURE 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: 294043; E313073;
 DR EMBL: 299121; E1186097;
 DR PFAM: PF00005; ABC_tran.1.
 KM HYPOTHETICAL PROTEIN.
 FT NON_TER 1 1
 SQ SEQUENCE 301 AA; 33779 MW; C3824B26 CRC32;
 Query Match 1.6%; Score 8; DB 2; Length 301;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 21 LKVDLE 28
 91 LKVDLE 98
 QY

RESULT 6
 ID 002155 PRELIMINARY; PRT; 327 AA.
 AC 002155;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CODED FOR BY C. ELEGANS CDNA YK77B5.5.
 GN T0984.5.
 OS CAENORABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITIA; RHABDITIOIDEA; RHABDITIDAE; PELODERINAE; CAENORABDITIDS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BRKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER S., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,


```

970 6 1.2 2195 3 002822 VESICLE COAT PROTEIN S 2.08e+02
971 6 1.2 2244 14 089344 POLYMERASE. 2.08e+02
972 6 1.2 2272 5 017329 GAG, POL AND ENV PROTE 2.08e+02
973 6 1.2 2304 14 088893 RNA1 POLYPROTEIN. 2.08e+02
974 6 1.2 2390 4 015020 BETA-SPECPRIN III. 2.08e+02
975 6 1.2 2478 13 093406 FIBRONECTIN. 2.08e+02
976 6 1.2 2489 3 006116 CHROMOSOME XVI COSMID 2.08e+02
977 6 1.2 2515 5 024551 TENASCIN-LIKE PROTEIN. 2.08e+02
978 6 1.2 2529 5 024605 ZINC FINGER PROTEIN. 2.08e+02
979 6 1.2 2606 5 021920 RILAB. 7 PROTEIN. 2.08e+02
980 6 1.2 2731 5 018366 ODD O2 PROTEIN. 2.08e+02
981 6 1.2 2731 5 061307 ODD O2 PROTEIN. 2.08e+02
982 6 1.2 2761 5 018447 CODED FOR BY C. ELEGAN 2.08e+02
983 6 1.2 2893 2 025063 TOXIN-LIKE OUTER MEMBR 2.08e+02
984 6 1.2 3019 14 068801 POLYPROTEIN. 2.08e+02
985 6 1.2 3063 2 059497 FATTY-ACID SYNTHASE (E 2.08e+02
986 6 1.2 3066 11 062388 ATAXIA TELANGIECTASIA 2.08e+02
987 6 1.2 3175 14 089939 REPLICASE ORF1B POLYPR 2.08e+02
988 6 1.2 3413 2 054593 POLYKETIDE SYNTHASE. 2.08e+02
989 6 1.2 3722 2 094873 ALPHA-AMINOADIPYL-CYST 2.08e+02
990 6 1.2 3972 2 073139 HYPOTHETICAL 418.3 KD 2.08e+02
991 6 1.2 4470 14 039225 RNA-DIRECTED RNA POLYM 2.08e+02
992 6 1.2 4560 4 013787 APOLIPOPROTEIN B100. 2.08e+02
993 6 1.2 4590 4 014517 CADHERIN-RELATED TUMOR 2.08e+02
994 6 1.2 4735 2 054666 POLYKETIDE SYNTHASE. 2.08e+02
995 6 1.2 4861 4 015751 P619. 2.08e+02
996 6 1.2 5069 2 052789 RIFAMYCIN POLYKETIDE S 2.08e+02
997 6 1.2 5071 5 091905 RYANODINE RECEPTOR. 2.08e+02
998 6 1.2 5107 5 094279 PARTIAL CDS. 2.08e+02
999 6 1.2 6420 5 095814 FK506 POLYKETIDE SYNTH 2.08e+02
1000 6 1.2 7829 5 018559 SIMILAR TO POLYKETIDE

```

ALIGNMENTS

PRELIMINARY: PRT: 495 AA.

```

RESULT 1
AC 002229; PRELIMINARY: PRT: 495 AA.
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE C48D1.2 PROTEIN.
GN C48D1.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BURTON J.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
KA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
KA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
KA CRAXTON M., DEAR S., DU Z., DUBBIN R., FAYELLO A., FULTON L.,
KA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,
KA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
KA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
KA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
KA SPALDON N., SMITH A., SONNHAMMER E., STADEN R., STUSTON J.,
KA THIERRY-MIES J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
KA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., MOHLMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
DR EMBL: 281049; E1344743;
DR PROSITE: PS01121; CASPASE_HIS: 1.
DR PROSITE: PS01122; CASPASE_CYS: 1.
SQ SEQUENCE 495 AA; 55549 MW; 7DC9ED9D CRC32;

```

Query Match 82.7%; Score 416; DB 5; Length 495;
Best Local Similarity 100.0%; Pred. No. 0.00e-00;

```

Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 MRODRSLLEINIMFSSHLKAYDELLEVLAKOVLNSNGMINSKGVRRKREIYAV 60
2 MRODRSLLEINIMFSSHLKAYDELLEVLAKOVLNSNGMINSKGVRRKREIYAV 61
61 ORRGVAFPAFYDALRSYGHEGLAEVLEPLARVSDSNVVECEPSPASHRSRALSPAG 120
62 ORRGVAFPAFYDALRSYGHEGLAEVLEPLARVSDSNVVECEPSPASHRSRALSPAG 121
121 YTSPTFRVARDVSYSVSTSYODIYSRARSRSRSHLSORHNSPVPNAFSPQSSA 180
122 YTSPTFRVARDVSYSVSTSYODIYSRARSRSRSHLSORHNSPVPNAFSPQSSA 181
181 NSSTFGCCSLGYSRRNSRFSKASPTQYIFHEEDMNVDAPTSRVDEKTMRNFSPP 240
182 NSSTFGCCSLGYSRRNSRFSKASPTQYIFHEEDMNVDAPTSRVDEKTMRNFSPP 241
241 RGMCLINNEHFEOMPTNRGTAKADNLTNLFRCMGTYVICKDNLTGMLLTIRDFAKH 300
242 RGMCLINNEHFEOMPTNRGTAKADNLTNLFRCMGTYVICKDNLTGMLLTIRDFAKH 301
301 ESHGDSAILVILSHGEENVYIGVDIPISTHEIYDLNANAPRLANKYIVQACGE 360
302 ESHGDSAILVILSHGEENVYIGVDIPISTHEIYDLNANAPRLANKYIVQACGE 361
361 RRNGFPVLDSDVGAFARFGMDNDRGDFENFGCVROVOVWRKKRPSADILI 416
362 RRNGFPVLDSDVGAFARFGMDNDRGDFENFGCVROVOVWRKKRPSADILI 417

```

PRELIMINARY: PRT: 102 AA.

```

RESULT 2
AC 079747; PRELIMINARY: PRT: 102 AA.
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYTOCHROME B (FRAGMENT).
GN OMANOSURA JAVAKARI.
OS MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; LEPIDOSAURIA; SQUAMATA;
OC SCLEROGLOSSA; SCINCOMORPHA; LACERTOIDEA; LACERTIDAE; OMANOSAURA.
RN [1]
RP SEQUENCE FROM N.A.
RA HARRIS D.J.;
RL "Relationships of lacertid lizards (Reptilia: Lacertidae) estimated
RL from mitochondrial DNA sequences and morphology."
DR EMBL: AF080351; G3560990;
KW MITOCHONDRION.
FT NON_TER 1
FT NON_TER 102
SQ SEQUENCE 102 AA; 11126 MW; 5B95367C CRC32;

```

Query Match 1.6%; Score 8; DB 8; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.37e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 3 SPRGMCLI 10
QY 240 SPRGMCLI 247

```

RESULT 3
AC 074357; PRELIMINARY: PRT: 174 AA.
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POSSIBLE SPLICED FACTOR (FRAGMENT).
GN SPK25D12.07C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;

824	6	1.2	845 14	036326	ENVELOPE PROTEIN.	2.08e+02	897	6	1.2	1028 2	068083	POTENTIAL OUTER MEMBR	2.08e+02
825	6	1.2	845 14	036330	ENVELOPE PROTEIN.	2.08e+02	898	6	1.2	1053 5	024310	POLYPROTEIN.	2.08e+02
826	6	1.2	845 14	036329	ENVELOPE PROTEIN.	2.08e+02	899	6	1.2	1054 5	078109	MHC CLASS II TRANSACTI	2.08e+02
827	6	1.2	845 14	036315	ENVELOPE PROTEIN.	2.08e+02	900	6	1.2	1072 11	035482	HIGH MOLECULAR-WEIGHT	2.08e+02
828	6	1.2	845 14	036318	ENVELOPE PROTEIN.	2.08e+02	901	6	1.2	1086 4	013423	NICOTINAMIDE NUCLEOTID	2.08e+02
829	6	1.2	845 14	036319	ENVELOPE PROTEIN.	2.08e+02	902	6	1.2	1086 4	016796	NAD(P) TRANSDHROGENAS	2.08e+02
830	6	1.2	845 14	036319	ENVELOPE PROTEIN.	2.08e+02	903	6	1.2	1094 14	088680	PURATIVE VIRAL RESPLICA	2.08e+02
831	6	1.2	845 14	036331	ENVELOPE PROTEIN.	2.08e+02	904	6	1.2	1095 14	060300	KIAA0553 PROTEIN (FRAG	2.08e+02
832	6	1.2	845 14	036338	ENVELOPE PROTEIN.	2.08e+02	905	6	1.2	1114 10	049345	PURATIVE RNA HELICASE	2.08e+02
833	6	1.2	845 14	036337	ENVELOPE PROTEIN.	2.08e+02	906	6	1.2	1116 14	098632	RNA BINDING PROTEIN.	2.08e+02
834	6	1.2	845 14	042056	ENVELOPE PROTEIN.	2.08e+02	907	6	1.2	1144 5	062179	F2H17.9 PROTEIN.	2.08e+02
835	6	1.2	845 14	042056	ENVELOPE PROTEIN.	2.08e+02	908	6	1.2	1148 14	055519	127K MINOR OUTER CASP	2.08e+02
836	6	1.2	846 5	017897	C10C5.1 PROTEIN.	2.08e+02	909	6	1.2	1152 5	045463	F36D3.5 PROTEIN.	2.08e+02
837	6	1.2	849 14	066731	ENVELOPE PROTEIN (FRAG	2.08e+02	910	6	1.2	1155 3	014239	HYPOHETICAL 128.3 KD	2.08e+02
838	6	1.2	852 11	088943	POTASSIUM CHANNEL.	2.08e+02	911	6	1.2	1157 5	027088	PHYRIVATE:FERREDOXIN OX	2.08e+02
839	6	1.2	853 14	089469	NEURONAL DELAYED-RECTI	2.08e+02	912	6	1.2	1157 5	027088	PHYRIVATE:FERREDOXIN OX	2.08e+02
840	6	1.2	854 4	075580	STEL2ALPHA.	2.08e+02	913	6	1.2	1163 4	092618	MYELOBLAST KIAA0222.	2.08e+02
841	6	1.2	855 3	042722	ENVELOPE POLYPROTEIN.	2.08e+02	914	6	1.2	1170 5	016587	F21E9.1 PROTEIN.	2.08e+02
842	6	1.2	858 14	089473	DNA GYRASE A SUBUNIT.	2.08e+02	915	6	1.2	1173 3	013313	MGOSIN-LIKE PROTEIN SP	2.08e+02
843	6	1.2	859 14	036321	ENVELOPE PROTEIN.	2.08e+02	916	6	1.2	1174 6	095168	TIGHT JUNCTION PROTEIN	2.08e+02
844	6	1.2	859 14	036325	ENVELOPE PROTEIN.	2.08e+02	917	6	1.2	1182 11	035595	PITCH2.	2.08e+02
845	6	1.2	859 14	036325	ENVELOPE PROTEIN.	2.08e+02	918	6	1.2	1186 5	062302	K04G2.8A PROTEIN.	2.08e+02
846	6	1.2	859 14	036324	ENVELOPE PROTEIN.	2.08e+02	919	6	1.2	1188 5	021227	K04G2.8B PROTEIN.	2.08e+02
847	6	1.2	859 14	036324	ENVELOPE PROTEIN.	2.08e+02	920	6	1.2	1192 5	061853	F33E11.3 PROTEIN.	2.08e+02
848	6	1.2	859 14	036350	ENVELOPE PROTEIN.	2.08e+02	921	6	1.2	1195 1	058524	HYPOHETICAL PROTEIN M	2.08e+02
849	6	1.2	859 14	036351	ENVELOPE PROTEIN.	2.08e+02	922	6	1.2	1198 2	037330	POLYKETIDE SYNTHASE (F	2.08e+02
850	6	1.2	859 14	036316	ENVELOPE PROTEIN.	2.08e+02	923	6	1.2	1199 5	091349	SIMILARITY TO SHORT RE	2.08e+02
851	6	1.2	859 14	092814	ENVELOPE POLYPROTEIN.	2.08e+02	924	6	1.2	1205 14	063300	IE-A PROTEIN.	2.08e+02
852	6	1.2	859 14	036356	ENVELOPE PROTEIN.	2.08e+02	925	6	1.2	1213 3	059801	PURATIVE HELICASE.	2.08e+02
853	6	1.2	859 14	036320	ENVELOPE PROTEIN.	2.08e+02	926	6	1.2	1214 5	025338	DELTA-LATROINSECTOXI	2.08e+02
854	6	1.2	859 14	036322	ENVELOPE PROTEIN.	2.08e+02	927	6	1.2	1232 1	058318	HYPOHETICAL PROTEIN M	2.08e+02
855	6	1.2	859 14	036334	ENVELOPE PROTEIN.	2.08e+02	928	6	1.2	1238 5	087830	SIMILARITY TO MOUSE SM	2.08e+02
856	6	1.2	859 14	036335	ENVELOPE PROTEIN.	2.08e+02	929	6	1.2	1259 8	035058	ATPA INTRON2 ORF.	2.08e+02
857	6	1.2	859 14	036336	ENVELOPE PROTEIN.	2.08e+02	930	6	1.2	1262 5	020771	F54D5.5.	2.08e+02
858	6	1.2	859 14	036337	ENVELOPE PROTEIN.	2.08e+02	931	6	1.2	1275 4	015057	KIAA0349 (FRAGMENT).	2.08e+02
859	6	1.2	859 14	036338	ENVELOPE PROTEIN.	2.08e+02	932	6	1.2	1277 11	035821	PAR INTERACTING PROTEI	2.08e+02
860	6	1.2	859 14	036353	ENVELOPE PROTEIN.	2.08e+02	933	6	1.2	1282 11	060520	PAIRED AMPHIPATHIC HEL	2.08e+02
861	6	1.2	859 14	036355	ENVELOPE PROTEIN.	2.08e+02	934	6	1.2	1287 3	006047	ANTIVIRAL PROTEIN (SWI	2.08e+02
862	6	1.2	859 14	036354	ENVELOPE PROTEIN.	2.08e+02	935	6	1.2	1291 10	038800	COL-O PUTATIVE RNA HEL	2.08e+02
863	6	1.2	859 14	042060	ENVELOPE PROTEIN.	2.08e+02	936	6	1.2	1301 14	077261	EG-115C2.10 PROTEIN.	2.08e+02
864	6	1.2	859 14	099268	ENV POLYPROTEIN PRECUR	2.08e+02	937	6	1.2	1302 14	065741	RNA-DIRECTED RNA POLYM	2.08e+02
865	6	1.2	859 14	036339	ENVELOPE PROTEIN.	2.08e+02	938	6	1.2	1315 3	074431	MEMBRANE ATPASE.	2.08e+02
866	6	1.2	859 14	021401	COSMID K09E3.	2.08e+02	939	6	1.2	1344 11	035851	P160 MYB-BINDING PROTE	2.08e+02
867	6	1.2	860 14	036333	ENVELOPE PROTEIN.	2.08e+02	940	6	1.2	1364 4	075092	MEF2 (FRAGMENT).	2.08e+02
868	6	1.2	867 5	044235	HRSH2.	2.08e+02	941	6	1.2	1371 2	073037	SENSOR TRANSDUCTION H	2.08e+02
869	6	1.2	868 10	038710	ABIETADYENE CYCLASE.	2.08e+02	942	6	1.2	1378 11	061138	PATERNALLY EXPRESSED P	2.08e+02
870	6	1.2	869 10	085041	CARBOXYTERME SHELL POLY	2.08e+02	943	6	1.2	1381 14	066628	CAPSID PROTEIN.	2.08e+02
871	6	1.2	871 5	017088	ZC308.1 PROTEIN.	2.08e+02	944	6	1.2	1463 14	099298	MAJOR IMMEDIATE EARLY	2.08e+02
872	6	1.2	880 5	013895	ATP-BINDING PROTEIN.	2.08e+02	945	6	1.2	1482 4	092509	MYELOBLAST KIAA0234.	2.08e+02
873	6	1.2	884 3	006163	CHROMOSOME XII COSMID	2.08e+02	946	6	1.2	1503 13	073677	CYSTIC FIBROSIS TRANSM	2.08e+02
874	6	1.2	884 7	029675	MHC CLASS II TRANSACTI	2.08e+02	947	6	1.2	1509 11	061194	PHOSPHATIDYLINOSITOL 3	2.08e+02
875	6	1.2	897 11	070495	PLENTY-OF-PROLINES-101	2.08e+02	948	6	1.2	1510 5	022699	SIMILARITY TO MYOSIN H	2.08e+02
876	6	1.2	899 11	063527	RETINOBLASTOMA PROTEIN	2.08e+02	949	6	1.2	1515 3	006625	ANALOG TO GLYCOSYL DE	2.08e+02
877	6	1.2	902 8	035063	COXI INTRON1 ORF.	2.08e+02	950	6	1.2	1551 4	014160	KIAA0147 PROTEIN (FRAG	2.08e+02
878	6	1.2	902 2	007686	ORF A PROTEIN.	2.08e+02	951	6	1.2	1551 4	014160	POLYMERIZATION POTENTI	2.08e+02
879	6	1.2	912 14	090278	ENVELOPE PROTEIN.	2.08e+02	952	6	1.2	1560 11	097868	ZINC FINGER PROTEIN.	2.08e+02
880	6	1.2	915 5	020099	SIMILAR TO DILUTE MYOS	2.08e+02	953	6	1.2	1571 11	054978	COSG5.6 (FRAGMENT).	2.08e+02
881	6	1.2	943 4	014664	PSD-95/SAP90-ASSOCIATE	2.08e+02	954	6	1.2	1602 5	017679	F8A11.1 PROTEIN.	2.08e+02
882	6	1.2	946 2	070888	(BCTA)	2.08e+02	955	6	1.2	1635 5	017368	PHOSPHATIDYLINOSITOL 3	2.08e+02
883	6	1.2	957 10	019204	SIMILAR TO THROMBOSPON	2.08e+02	956	6	1.2	1658 11	061182	RNA POLYMERASE II.	2.08e+02
884	6	1.2	959 10	023291	KINESIN HOMOLOG.	2.08e+02	957	6	1.2	1661 5	027095	CENTROMERE PROTEIN HOM	2.08e+02
885	6	1.2	959 10	022987	UNKNOWN PROTEIN.	2.08e+02	958	6	1.2	1676 10	023332	PHOSPHOINOSITIDE 3-KIN	2.08e+02
886	6	1.2	965 5	027502	K03A11.1 PROTEIN.	2.08e+02	959	6	1.2	1686 4	000443	(CDNA1) PROTEIN 4.1 HO	2.08e+02
887	6	1.2	969 5	027502	SIMILARITY TO ARABIDOP	2.08e+02	960	6	1.2	1698 5	024440	PUTATIVE OUTER MEMBRAN	2.08e+02
888	6	1.2	979 5	053609	PUTATIVE MEMBRANE PROT	2.08e+02	961	6	1.2	1742 5	024463	POLYPROTEIN.	2.08e+02
889	6	1.2	994 5	077089	DORSAL B.	2.08e+02	962	6	1.2	1770 2	084419	HIM-2	2.08e+02
890	6	1.2	997 3	074325	HYPOHETICAL 111.3 KD	2.08e+02	963	6	1.2	1834 14	066233	T7123.15 PROTEIN.	2.08e+02
891	6	1.2	1001 11	088664	SERINE/THREONINE PROTE	2.08e+02	964	6	1.2	1865 10	081909	COSMID C06A5.	2.08e+02
892	6	1.2	1003 5	001915	SIMILARITY TO N-CHIME	2.08e+02	965	6	1.2	1932 5	001483	PKS MODULE 4.	2.08e+02
893	6	1.2	1005 1	058718	HYPOHETICAL PROTEIN M	2.08e+02	966	6	1.2	1937 2	030482	3D7VARI (FRAGMENT).	2.08e+02
894	6	1.2	1017 11	089048	ELK CHANNEL 1.	2.08e+02	967	6	1.2	2042 5	025766	NDMA PROTEIN.	2.08e+02
895	6	1.2	1026 3	006315	CHROMOSOME XII COSMID	2.08e+02	968	6	1.2	2115 4	014980	LE.MBI PROTEIN.	2.08e+02
896	6	1.2	1028 11	015042	KIAA0332 (FRAGMENT).	2.08e+02	969	6	1.2	2140 3	002316		

678	6	1.2	566 14	098095	HEMAGGLUTININ. PRECURS	2.08e+02	751	6	1.2	680 11	070583	MIDLINE 1 (MIDLINE 1 P	2.08e+02
679	6	1.2	566 14	082774	HEMAGGLUTININ. PRECURS	2.08e+02	752	6	1.2	684 1	058969	HYDROLYTIC HELICASE	2.08e+02
680	6	1.2	566 14	082503	HEMAGGLUTININ. HA (FRA	2.08e+02	753	6	1.2	689 13	073620	NUCLEAR PROTEIN SDRZ.	2.08e+02
681	6	1.2	566 14	088836	HEMAGGLUTININ. HA (FRA	2.08e+02	754	6	1.2	695 5	091908	MOA7.2 PROTEIN.	2.08e+02
682	6	1.2	566 14	016716	PYRUVATE KINASE (EC 2.	2.08e+02	755	6	1.2	697 14	087026	CAPSID.	2.08e+02
683	6	1.2	566 14	082770	HEMAGGLUTININ. PRECURSO	2.08e+02	756	6	1.2	699 5	019640	COSMID F20D12.	2.08e+02
684	6	1.2	566 14	021740	ROSD1.8 PROTEIN.	2.08e+02	757	6	1.2	700 13	091453	STONUSOXIN BETA-SUBUN	2.08e+02
685	6	1.2	566 14	082832	HEMAGGLUTININ. PRECURSO	2.08e+02	758	6	1.2	701 2	051403	NA/H+ ANTIporter (NAP	2.08e+02
686	6	1.2	566 14	084110	HEMAGGLUTININ.	2.08e+02	759	6	1.2	701 14	065568	HYDROLYTIC 72.6 KD P	2.08e+02
687	6	1.2	566 14	082500	HEMAGGLUTININ. PRECURSO	2.08e+02	760	6	1.2	703 6	079122	CHROMOSOME XV READING	2.08e+02
688	6	1.2	566 14	067043	HEMAGGLUTININ. PRECURSO	2.08e+02	761	6	1.2	706 3	012296	ORE H130.	2.08e+02
689	6	1.2	566 14	067042	HEMAGGLUTININ. PRECURSO	2.08e+02	762	6	1.2	706 10	081324	VERUCOTOXIN ALPHA PRE	2.08e+02
690	6	1.2	566 14	021728	SIMILAR TO CYTOPLASMIC	2.08e+02	763	6	1.2	708 13	098993	STIMULANT TO TPR DOMA	2.08e+02
691	6	1.2	566 14	021728	TRBL.	2.08e+02	764	6	1.2	708 13	022961	NUCLEAR PROTEIN SDR3 (2.08e+02
692	6	1.2	566 14	021728	TRBL.	2.08e+02	765	6	1.2	708 13	022961	WUGSC:H.D0808A01.2 PR	2.08e+02
693	6	1.2	566 14	021728	TRBL.	2.08e+02	766	6	1.2	708 13	022961	TRANSLATION INITIATION	2.08e+02
694	6	1.2	566 14	021728	TRBL.	2.08e+02	767	6	1.2	708 13	022961	PUTATIVE TRANSMEMBRANE	2.08e+02
695	6	1.2	566 14	021728	TRBL.	2.08e+02	768	6	1.2	708 13	022961	NUCLEAR PROTEIN.	2.08e+02
696	6	1.2	566 14	021728	TRBL.	2.08e+02	769	6	1.2	708 13	022961	ENVELOPE PROTEIN.	2.08e+02
697	6	1.2	566 14	021728	TRBL.	2.08e+02	770	6	1.2	708 13	022961	X-DELTA-1.	2.08e+02
698	6	1.2	566 14	021728	TRBL.	2.08e+02	771	6	1.2	708 13	022961	PININ.	2.08e+02
699	6	1.2	566 14	021728	TRBL.	2.08e+02	772	6	1.2	708 13	022961	PCF1P. HOMOLOG.	2.08e+02
700	6	1.2	566 14	021728	TRBL.	2.08e+02	773	6	1.2	708 13	022961	PROBABLE DNA TOPOISOMER	2.08e+02
701	6	1.2	566 14	021728	TRBL.	2.08e+02	774	6	1.2	708 13	022961	HYDROLYTIC 84.2 KD P	2.08e+02
702	6	1.2	566 14	021728	TRBL.	2.08e+02	775	6	1.2	708 13	022961	MAIATE SYNTHASE G.	2.08e+02
703	6	1.2	566 14	021728	TRBL.	2.08e+02	776	6	1.2	708 13	022961	GTP PYRROPHOSPHOKINASE	2.08e+02
704	6	1.2	566 14	021728	TRBL.	2.08e+02	777	6	1.2	708 13	022961	TOSALD.4 PROTEIN.	2.08e+02
705	6	1.2	566 14	021728	TRBL.	2.08e+02	778	6	1.2	708 13	022961	PLASMAID RP4 TRE.	2.08e+02
706	6	1.2	566 14	021728	TRBL.	2.08e+02	779	6	1.2	708 13	022961	P1067 PROTEIN.	2.08e+02
707	6	1.2	566 14	021728	TRBL.	2.08e+02	780	6	1.2	708 13	022961	VPI LIKE PROTEIN (FRAG	2.08e+02
708	6	1.2	566 14	021728	TRBL.	2.08e+02	781	6	1.2	708 13	022961	FAS BINDING PROTEIN (D	2.08e+02
709	6	1.2	566 14	021728	TRBL.	2.08e+02	782	6	1.2	708 13	022961	PININ.	2.08e+02
710	6	1.2	566 14	021728	TRBL.	2.08e+02	783	6	1.2	708 13	022961	W04D.6 PROTEIN.	2.08e+02
711	6	1.2	566 14	021728	TRBL.	2.08e+02	784	6	1.2	708 13	022961	1-EVIDENCE-PREDICTED B	2.08e+02
712	6	1.2	566 14	021728	TRBL.	2.08e+02	785	6	1.2	708 13	022961	OREF10.	2.08e+02
713	6	1.2	566 14	021728	TRBL.	2.08e+02	786	6	1.2	708 13	022961	MC032L.	2.08e+02
714	6	1.2	566 14	021728	TRBL.	2.08e+02	787	6	1.2	708 13	022961	SERINE/THREONINE PROTE	2.08e+02
715	6	1.2	566 14	021728	TRBL.	2.08e+02	788	6	1.2	708 13	022961	CARS-CYP (EC 5.2.1.8).	2.08e+02
716	6	1.2	566 14	021728	TRBL.	2.08e+02	789	6	1.2	708 13	022961	TIB KINASE BETA.	2.08e+02
717	6	1.2	566 14	021728	TRBL.	2.08e+02	790	6	1.2	708 13	022961	PROBABLE NA(+)/H(+) AN	2.08e+02
718	6	1.2	566 14	021728	TRBL.	2.08e+02	791	6	1.2	708 13	022961	BETA-GLUCOSIDASE.	2.08e+02
719	6	1.2	566 14	021728	TRBL.	2.08e+02	792	6	1.2	708 13	022961	F88E6.1 PROTEIN.	2.08e+02
720	6	1.2	566 14	021728	TRBL.	2.08e+02	793	6	1.2	708 13	022961	CHITINASE.	2.08e+02
721	6	1.2	566 14	021728	TRBL.	2.08e+02	794	6	1.2	708 13	022961	HYDROLYTIC 87.5 KD P	2.08e+02
722	6	1.2	566 14	021728	TRBL.	2.08e+02	795	6	1.2	708 13	022961	PININ.	2.08e+02
723	6	1.2	566 14	021728	TRBL.	2.08e+02	796	6	1.2	708 13	022961	PROBABLE SERINE/THREON	2.08e+02
724	6	1.2	566 14	021728	TRBL.	2.08e+02	797	6	1.2	708 13	022961	VP4 PROTEIN.	2.08e+02
725	6	1.2	566 14	021728	TRBL.	2.08e+02	798	6	1.2	708 13	022961	TEX PROTEIN (TEX).	2.08e+02
726	6	1.2	566 14	021728	TRBL.	2.08e+02	799	6	1.2	708 13	022961	CHROMOSOME XII COSMID	2.08e+02
727	6	1.2	566 14	021728	TRBL.	2.08e+02	800	6	1.2	708 13	022961	CHROMOSOME IV READING	2.08e+02
728	6	1.2	566 14	021728	TRBL.	2.08e+02	801	6	1.2	708 13	022961	F50D6.1 PROTEIN (FRAGM	2.08e+02
729	6	1.2	566 14	021728	TRBL.	2.08e+02	802	6	1.2	708 13	022961	NUCLEOLAR RNA HELICASE	2.08e+02
730	6	1.2	566 14	021728	TRBL.	2.08e+02	803	6	1.2	708 13	022961	HISTONE TRANSCRIPTOMA	2.08e+02
731	6	1.2	566 14	021728	TRBL.	2.08e+02	804	6	1.2	708 13	022961	UL47H.	2.08e+02
732	6	1.2	566 14	021728	TRBL.	2.08e+02	805	6	1.2	708 13	022961	HYDROLYTIC 91.3 KD P	2.08e+02
733	6	1.2	566 14	021728	TRBL.	2.08e+02	806	6	1.2	708 13	022961	T2H6.6 PROTEIN.	2.08e+02
734	6	1.2	566 14	021728	TRBL.	2.08e+02	807	6	1.2	708 13	022961	C2-HC TYPE ZINC FINGER	2.08e+02
735	6	1.2	566 14	021728	TRBL.	2.08e+02	808	6	1.2	708 13	022961	NK-TUMOR RECOGNITION M	2.08e+02
736	6	1.2	566 14	021728	TRBL.	2.08e+02	809	6	1.2	708 13	022961	CODED FOR BY C. ELISAN	2.08e+02
737	6	1.2	566 14	021728	TRBL.	2.08e+02	810	6	1.2	708 13	022961	B16A LONG HYPOTHETICA	2.08e+02
738	6	1.2	566 14	021728	TRBL.	2.08e+02	811	6	1.2	708 13	022961	P1074 PROTEIN.	2.08e+02
739	6	1.2	566 14	021728	TRBL.	2.08e+02	812	6	1.2	708 13	022961	SEF/ARG-RELATED NUCLEA	2.08e+02
740	6	1.2	566 14	021728	TRBL.	2.08e+02	813	6	1.2	708 13	022961	95.1KD PUTATIVE NONSTR	2.08e+02
741	6	1.2	566 14	021728	TRBL.	2.08e+02	814	6	1.2	708 13	022961	TAI195.	2.08e+02
742	6	1.2	566 14	021728	TRBL.	2.08e+02	815	6	1.2	708 13	022961	PUTATIVE ZINC METALLOP	2.08e+02
743	6	1.2	566 14	021728	TRBL.	2.08e+02	816	6	1.2	708 13	022961	CONDUCTIN.	2.08e+02
744	6	1.2	566 14	021728	TRBL.	2.08e+02	817	6	1.2	708 13	022961	RELA PROTEIN.	2.08e+02
745	6	1.2	566 14	021728	TRBL.	2.08e+02	818	6	1.2	708 13	022961	POLLEN-SPECIFIC PROTEI	2.08e+02
746	6	1.2	566 14	021728	TRBL.	2.08e+02	819	6	1.2	708 13	022961	VOLTAGE GATED POTASSIU	2.08e+02
747	6	1.2	566 14	021728	TRBL.	2.08e+02	820	6	1.2	708 13	022961		
748	6	1.2	566 14	021728	TRBL.	2.08e+02	821	6	1.2	708 13	022961		
749	6	1.2	566 14	021728	TRBL.	2.08e+02	822	6	1.2	708 13	022961		
750	6	1.2	566 14	021728	TRBL.	2.08e+02	823	6	1.2	708 13	022961		

532	6	1.2	401	5	045907	NHR-65 PROTEIN.	2.08e+02	605	6	1.2	483	14	041517	55 KDA PROTEIN.	2.08e+02
533	6	1.2	404	13	090420	47 KDA HEAT SHOCK PROT	2.08e+02	606	6	1.2	487	2	050901	BETA-1,4-GLUCANASE.	2.08e+02
534	6	1.2	407	13	P70053	PAX-6.	2.08e+02	607	6	1.2	489	10	065518	HYPOTHETICAL 54.1 KD P	2.08e+02
535	6	1.2	408	2	007796	POSSIBLE TRANSPOSASE.	2.08e+02	608	6	1.2	490	2	P95098	HYPOTHETICAL 54.0 KD P	2.08e+02
537	6	1.2	409	3	012128	YOR3220W FROM CHROMOSO	2.08e+02	609	6	1.2	491	10	049660	PREDICTED PROTEIN.	2.08e+02
539	6	1.2	410	13	090323	SERINE PROTEASE INHIBI	2.08e+02	610	6	1.2	492	13	093352	OTOKERATIN.	2.08e+02
540	6	1.2	412	11	063556	SERINE PROTEASE INHIBI	2.08e+02	611	6	1.2	494	5	P91653	SERENDIPITY (FRAGMENT)	2.08e+02
541	6	1.2	413	5	062467	Y45F108.2 PROTEIN.	2.08e+02	612	6	1.2	496	4	015414	YRM1 PROTEIN.	2.08e+02
544	6	1.2	416	2	061734	BRAIN AND MUSCLE ARNT-	2.08e+02	613	6	1.2	496	4	025600	CONSERVED HYPOTHETICAL	2.08e+02
545	6	1.2	417	2	025751	HYPOTHETICAL 47.8 KD P	2.08e+02	614	6	1.2	497	5	P91708	SERENDIPITY (FRAGMENT)	2.08e+02
548	6	1.2	417	2	025751	COLITIN TOLERANCE-LIKE	2.08e+02	615	6	1.2	498	5	P91709	SENDERPITY (FRAGMENT)	2.08e+02
549	6	1.2	417	11	035329	PRODUCT OF THE 2.2KB T	2.08e+02	616	6	1.2	499	4	013675	ALPHA 1C ADRENERGIC RE	2.08e+02
550	6	1.2	417	11	035329	IPL1 AND AURORA RELATE	2.08e+02	617	6	1.2	500	11	088962	STEROL 12-ALPHA HYDROX	2.08e+02
551	6	1.2	418	2	025454	CONSERVED HYPOTHETICAL	2.08e+02	618	6	1.2	500	6	002766	STEROL 12-ALPHA HYDROX	2.08e+02
552	6	1.2	418	4	099413	PAX6 (FRAGMENT).	2.08e+02	619	6	1.2	501	11	035074	POSTFACCYLIN SYNTHASE	2.08e+02
553	6	1.2	420	14	067008	POLYPROTEIN PRECURSOR	2.08e+02	620	6	1.2	501	11	075004	CDC24P.	2.08e+02
554	6	1.2	420	14	067008	HYPOTHETICAL 46.1 KD P	2.08e+02	621	6	1.2	501	11	062969	PROSTACYCLIN SYNTHASE.	2.08e+02
555	6	1.2	421	13	P70001	XENOPUS PAX-6 LONG (PR	2.08e+02	622	6	1.2	505	10	064667	F22013.18.	2.08e+02
556	6	1.2	421	10	034609	PROTEIN KINASE - LIKE	2.08e+02	623	6	1.2	509	8	047547	FLTA22.35 PROTEIN.	2.08e+02
557	6	1.2	422	11	035669	FLICE PARTIAL CDS (FLI	2.08e+02	624	6	1.2	509	10	040870	CYCLOCHROME C OXIDASE S	2.08e+02
558	6	1.2	422	2	025122	HRO-CDC25.	2.08e+02	625	6	1.2	510	2	059913	LEGUMIN-LIKE STORAGE P	2.08e+02
559	6	1.2	423	10	082626	HYPOTHETICAL 44.1 KD P	2.08e+02	626	6	1.2	510	2	059913	RNA POLYMERASE SIGMA F	2.08e+02
560	6	1.2	423	10	082626	ERG PROTEIN.	2.08e+02	627	6	1.2	511	14	011450	ENVELOPE GLYCOPROTEIN.	2.08e+02
561	6	1.2	424	1	026914	CONSERVED PROTEIN.	2.08e+02	628	6	1.2	511	5	016512	ENVELOPE GLYCOPROTEIN.	2.08e+02
562	6	1.2	425	1	025142	FUCOSYLTRANSFERASE.	2.08e+02	629	6	1.2	513	5	016293	T05H4.7 PROTEIN.	2.08e+02
563	6	1.2	425	11	070290	G-PROTEIN-COUPLED INMA	2.08e+02	630	6	1.2	513	2	067052	F32D1.9 PROTEIN.	2.08e+02
564	6	1.2	427	4	000148	NUCLEAR RNA HELICASE.	2.08e+02	631	6	1.2	515	14	093015	HISTIDINE KINASE SENSO	2.08e+02
565	6	1.2	427	5	023491	COSMID T14P9.	2.08e+02	632	6	1.2	515	3	099229	REQUIRE FOR VACUOLE S	2.08e+02
566	6	1.2	428	2	033959	TYL ORFX.	2.08e+02	633	6	1.2	517	10	065519	HYPOTHETICAL 58.5 KD P	2.08e+02
567	6	1.2	429	4	017729	ALPHA 1C ADRENERGIC RE	2.08e+02	634	6	1.2	520	5	093877	KO9E9.1 PROTEIN.	2.08e+02
568	6	1.2	432	14	009639	HEMAGGLUTININ PRECURSO	2.08e+02	635	6	1.2	521	10	022848	MEMBRANE TRANSPORTER D	2.08e+02
569	6	1.2	433	2	051871	POLY(3-HYDROXYBUTYRATE	2.08e+02	636	6	1.2	521	10	048926	CYP93C1P.	2.08e+02
570	6	1.2	436	13	042348	PAX6 PROTEIN.	2.08e+02	637	6	1.2	521	5	076360	F55A8.2 PROTEIN (FRAGM	2.08e+02
571	6	1.2	437	10	009638	HEMAGGLUTININ PRECURSO	2.08e+02	638	6	1.2	523	4	099332	HSP-1.	2.08e+02
572	6	1.2	440	2	064810	PUTATIVE SERINE CARBOX	2.08e+02	639	6	1.2	528	2	092483	RNA POLYMERASE SIGMA F	2.08e+02
573	6	1.2	441	14	007499	HYPOTHETICAL 47.7 KD P	2.08e+02	640	6	1.2	530	14	092483	GP64/67-EF-ACNPV ORF	2.08e+02
574	6	1.2	441	14	007499	NUCLEOCAPSID PROTEIN.	2.08e+02	641	6	1.2	530	5	018064	C.	2.08e+02
575	6	1.2	442	10	023335	RNA HELICASE HOMOLOG.	2.08e+02	642	6	1.2	531	2	055524	HEMAGGLUTININ PRECURSO	2.08e+02
576	6	1.2	443	2	087843	HYPOTHETICAL 48.1 KD P	2.08e+02	643	6	1.2	531	2	055524	HYPOTHETICAL 60.6 KD P	2.08e+02
577	6	1.2	445	10	080503	F16H22.13 PROTEIN.	2.08e+02	644	6	1.2	532	5	017828	F21A3.5 PROTEIN.	2.08e+02
578	6	1.2	451	14	092492	MES3-ACMNPV ORF139.	2.08e+02	645	6	1.2	533	11	055221	PUTATIVE CD98 PROTEIN.	2.08e+02
579	6	1.2	451	5	020648	CODED FOR BY C. ELEGAN	2.08e+02	646	6	1.2	533	5	043989	HOMOBBOX-CONTAINING PR	2.08e+02
580	6	1.2	451	5	019321	RIBULOSE BIPHOSPHATE	2.08e+02	647	6	1.2	534	14	066643	DNA HELICASE-PRIMASE C	2.08e+02
581	6	1.2	451	8	032328	TYROSINE KINASE.	2.08e+02	648	6	1.2	535	3	042653	HYPOTHETICAL 62.0 KD P	2.08e+02
582	6	1.2	453	5	P92013	R10D12.15 PROTEIN.	2.08e+02	649	6	1.2	535	10	049600	IMPORTIN ALPHA-LIKE PR	2.08e+02
583	6	1.2	453	5	022803	PAIRED-TYPE HOMEDOMAI	2.08e+02	650	6	1.2	538	11	088900	MOLECULAR ADAPTER RGRB	2.08e+02
584	6	1.2	454	4	014060	COPROPHETIRINOGEN OXI	2.08e+02	651	6	1.2	539	10	013971	CDG25 B-TYPE TYROSINE	2.08e+02
585	6	1.2	455	4	060451	ALPHA 1A ADRENERGIC RE	2.08e+02	652	6	1.2	539	5	081053	T18E12.12 PROTEIN.	2.08e+02
586	6	1.2	459	10	022803	PUTATIVE SERINE CARBOX	2.08e+02	653	6	1.2	543	5	021064	CYTOSOLSMIC INTERMEDIA	2.08e+02
587	6	1.2	462	2	P95644	OUTER MEMBRANE PROTEIN	2.08e+02	654	6	1.2	543	5	012866	SIMILAR TO INTERMEDIAT	2.08e+02
588	6	1.2	462	2	P95644	RNA POLYMERASE SIGMA F	2.08e+02	655	6	1.2	546	3	008904	CHROMOSOME XV READING	2.08e+02
589	6	1.2	462	4	075958	STEROL 12-ALPHA HYDROX	2.08e+02	656	6	1.2	549	5	019287	SIMILAR TO INTERMEDIAT	2.08e+02
590	6	1.2	464	4	014676	APOPTOTIC CASPASE MCHS	2.08e+02	657	6	1.2	552	10	096336	AMP-BINDING PROTEIN.	2.08e+02
591	6	1.2	466	11	054913	RNA POLYMERASE BETA' S	2.08e+02	658	6	1.2	553	1	026576	PIRVINATE DEHYDROENASE	2.08e+02
592	6	1.2	467	3	036014	ALPHA 1A-ADRENERGIC RE	2.08e+02	659	6	1.2	556	4	081808	HYPOTHETICAL 62.6 KD P	2.08e+02
593	6	1.2	467	3	074430	PROBABLE VACUOLAR AMIN	2.08e+02	660	6	1.2	556	4	015530	3-PHOSPHOINOSITIDE DEP	2.08e+02
594	6	1.2	470	2	P95121	TRIGLYCERIDE LIPASE-CH	2.08e+02	661	6	1.2	559	11	054840	T-BOX 14 (NM0B14).	2.08e+02
595	6	1.2	471	5	P90988	HYPOTHETICAL 46.7 KD P	2.08e+02	662	6	1.2	559	11	054840	PKB KINASE.	2.08e+02
596	6	1.2	471	1	028786	COSMID B0432.	2.08e+02	663	6	1.2	561	14	082764	HEMAGGLUTININ (FRAGMEN	2.08e+02
597	6	1.2	471	1	028786	HYPOTHETICAL 54.2 KD P	2.08e+02	664	6	1.2	561	14	082764	HEMAGGLUTININ (FRAGMEN	2.08e+02
598	6	1.2	472	1	020178	PARACRYSTALLINE SURFAC	2.08e+02	665	6	1.2	562	14	082764	HEMAGGLUTININ PRECURSO	2.08e+02
599	6	1.2	472	1	020178	PARACRYSTALLINE SURFAC	2.08e+02	666	6	1.2	562	2	086284	PUTATIVE IMMUNITY.	2.08e+02
600	6	1.2	476	2	025366	FUCOSYLTRANSFERASE.	2.08e+02	667	6	1.2	563	10	065520	HYPOHETICAL 62.5 KD P	2.08e+02
601	6	1.2	477	5	026218	CYCLOCHROME C OXIDASE S	2.08e+02	668	6	1.2	565	10	049520	HYPOHETICAL 63.6 KD P	2.08e+02
602	6	1.2	477	8	063695	CYCLOCHROME C OXIDASE S	2.08e+02	669	6	1.2	565	14	082766	HEMAGGLUTININ PRECURSO	2.08e+02
603	6	1.2	478	5	025684	MAJOR MEMOZOITE SURFAC	2.08e+02	670	6	1.2	565	10	048774	CARROT B2 PROTEIN-LIKE	2.08e+02
604	6	1.2	478	11	062905	VITRONECTIN.	2.08e+02	671	6	1.2	566	14	067039	HEMAGGLUTININ.	2.08e+02
605	6	1.2	478	11	062905	ALPHA1,3-FUCOSYLTRANSF	2.08e+02	672	6	1.2	566	14	067040	HEMAGGLUTININ.	2.08e+02
606	6	1.2	480	11	089110	CASPASE-8.	2.08e+02	673	6	1.2	566	14	082834	HEMAGGLUTININ PRECURSO	2.08e+02
607	6	1.2	480	11	089110	PS6D6.2 PROTEIN.	2.08e+02	674	6	1.2	566	14	067007	POLYPROTEIN PRECURSOR.	2.08e+02
608	6	1.2	480	5	061159	D2089.1 PROTEIN.	2.08e+02	675	6	1.2	566	14	098092	HEMAGGLUTININ.	2.08e+02
609	6	1.2	480	5	061159	D2089.1 PROTEIN.	2.08e+02	676	6	1.2	566	14	098094	HEMAGGLUTININ.	2.08e+02
610	6	1.2	480	5	061159	D2089.1 PROTEIN.	2.08e+02	677	6	1.2	566	14	098093	HEMAGGLUTININ.	2.08e+02

386	6	1.2	344 14	007767	HEMAGGLUTININ PRECURSO	2.08e+02	459	6	1.2	361 7	046881	MHC CLASS I DLA-64.	2.08e+02
387	6	1.2	344 14	007753	HEMAGGLUTININ PRECURSO	2.08e+02	460	6	1.2	364 14	068800	NS5 (FRAGMENT).	2.08e+02
388	6	1.2	344 14	007774	HEMAGGLUTININ PRECURSO	2.08e+02	461	6	1.2	364 14	068816	NS5 (FRAGMENT).	2.08e+02
389	6	1.2	344 14	067060	HEMAGGLUTININ (FRAGME	2.08e+02	462	6	1.2	364 14	068795	NS5 (FRAGMENT).	2.08e+02
390	6	1.2	344 14	067080	HEMAGGLUTININ (FRAGME	2.08e+02	463	6	1.2	364 14	068755	NS5 (FRAGMENT).	2.08e+02
391	6	1.2	344 14	067066	HEMAGGLUTININ (FRAGME	2.08e+02	464	6	1.2	364 14	068815	NS5 (FRAGMENT).	2.08e+02
392	6	1.2	344 14	067066	HEMAGGLUTININ (FRAGME	2.08e+02	465	6	1.2	364 14	081553	NS5 (FRAGMENT).	2.08e+02
393	6	1.2	344 14	067082	HEMAGGLUTININ (FRAGME	2.08e+02	466	6	1.2	364 14	068789	NS5 (FRAGMENT).	2.08e+02
394	6	1.2	344 14	067077	HEMAGGLUTININ (FRAGME	2.08e+02	467	6	1.2	364 14	068809	NS5 (FRAGMENT).	2.08e+02
395	6	1.2	344 14	067075	HEMAGGLUTININ (FRAGME	2.08e+02	468	6	1.2	364 14	068807	NS5 (FRAGMENT).	2.08e+02
396	6	1.2	344 14	010623	HEMAGGLUTININ (FRAGME	2.08e+02	469	6	1.2	364 14	077572	X-LINKED VISUAL PIGMEN	2.08e+02
397	6	1.2	344 14	067109	HEMAGGLUTININ PRECURS	2.08e+02	470	6	1.2	364 6	077576	X-LINKED VISUAL PIGMEN	2.08e+02
398	6	1.2	344 14	067069	HEMAGGLUTININ (FRAGME	2.08e+02	471	6	1.2	364 6	077575	X-LINKED VISUAL PIGMEN	2.08e+02
399	6	1.2	344 14	067083	HEMAGGLUTININ (FRAGME	2.08e+02	472	6	1.2	364 2	044471	PUTATIVE TARTRATE DEHY	2.08e+02
400	6	1.2	344 14	067084	HEMAGGLUTININ PRECURS	2.08e+02	473	6	1.2	364 6	062860	LONG-WAVELENGTH SENSIT	2.08e+02
401	6	1.2	344 14	067129	HEMAGGLUTININ PRECURS	2.08e+02	474	6	1.2	364 6	077571	COLOR VISION OPSIN.	2.08e+02
402	6	1.2	344 14	067100	HEMAGGLUTININ PRECURS	2.08e+02	475	6	1.2	364 6	077570	X-LINKED VISUAL PIGMEN	2.08e+02
403	6	1.2	344 14	067096	HEMAGGLUTININ PRECURS	2.08e+02	476	6	1.2	364 6	077573	X-LINKED VISUAL PIGMEN	2.08e+02
404	6	1.2	344 14	067114	HEMAGGLUTININ PRECURS	2.08e+02	477	6	1.2	364 6	077574	X-LINKED VISUAL PIGMEN	2.08e+02
405	6	1.2	344 14	067124	HEMAGGLUTININ PRECURS	2.08e+02	478	6	1.2	364 6	077574	X-LINKED VISUAL PIGMEN	2.08e+02
406	6	1.2	344 14	067121	HEMAGGLUTININ PRECURS	2.08e+02	479	6	1.2	366 5	016891	F13A2.6 PROTEIN.	2.08e+02
407	6	1.2	344 14	067076	HEMAGGLUTININ (FRAGME	2.08e+02	480	6	1.2	367 14	041165	A683L PROTEIN.	2.08e+02
408	6	1.2	344 14	067130	HEMAGGLUTININ PRECURS	2.08e+02	481	6	1.2	368 11	097353	GDP-L-FUCOSE: BETA-D-G	2.08e+02
409	6	1.2	344 14	067068	HEMAGGLUTININ (FRAGME	2.08e+02	482	6	1.2	369 11	069143	PUTATIVE PYRUVATE-FERR	2.08e+02
410	6	1.2	344 14	067067	HEMAGGLUTININ (FRAGME	2.08e+02	483	6	1.2	370 13	P70002	XENOPUS PAX-6 SHORT (F	2.08e+02
411	6	1.2	344 14	067061	HEMAGGLUTININ (FRAGME	2.08e+02	484	6	1.2	370 10	022840	UNKNOWN PROTEIN.	2.08e+02
412	6	1.2	344 14	067101	HEMAGGLUTININ PRECURS	2.08e+02	485	6	1.2	370 10	009637	HEMAGGLUTININ PRECURSO	2.08e+02
413	6	1.2	344 14	067127	HEMAGGLUTININ PRECURS	2.08e+02	486	6	1.2	370 10	009636	HEMAGGLUTININ PRECURSO	2.08e+02
414	6	1.2	344 14	067065	HEMAGGLUTININ (FRAGME	2.08e+02	487	6	1.2	371 10	081734	HYPOHETICAL 40.8 KD P	2.08e+02
415	6	1.2	344 14	007776	HEMAGGLUTININ PRECURSO	2.08e+02	488	6	1.2	371 2	085595	ALANINE DEHYDROGENASE	2.08e+02
416	6	1.2	344 14	007781	HEMAGGLUTININ PRECURSO	2.08e+02	489	6	1.2	371 2	075381	PROXISOMAL MEMBRANE A	2.08e+02
417	6	1.2	344 14	007780	HEMAGGLUTININ PRECURSO	2.08e+02	490	6	1.2	377 4	043534	NATURAL KILLER CELL IN	2.08e+02
418	6	1.2	344 14	007770	HEMAGGLUTININ PRECURSO	2.08e+02	491	6	1.2	377 4	095706	NATURAL KILLER CELL RE	2.08e+02
419	6	1.2	344 14	007775	HEMAGGLUTININ PRECURSO	2.08e+02	492	6	1.2	377 4	095600	NK RECEPTOR.	2.08e+02
420	6	1.2	344 14	007775	HEMAGGLUTININ PRECURSO	2.08e+02	493	6	1.2	377 4	014621	KILLER CELL RECEPTOR.	2.08e+02
421	6	1.2	347 3	059756	MHC CLASS I ANTIGEN (F	2.08e+02	494	6	1.2	377 4	075601	EVOLUTIONARILY RELATED	2.08e+02
422	6	1.2	347 3	059756	HYPOHETICAL 39.0 KD P	2.08e+02	495	6	1.2	377 4	029504	HYPOHETICAL 42.7 KD P	2.08e+02
423	6	1.2	349 5	021602	UV-SENSITIVE CONE OPSI	2.08e+02	496	6	1.2	379 8	048364	CYTCHROME B.	2.08e+02
424	6	1.2	349 5	060399	M88.4 PROTEIN.	2.08e+02	497	6	1.2	379 8	047992	CYTCHROME B.	2.08e+02
425	6	1.2	350 4	007564	CGM6-HUMAN.	2.08e+02	498	6	1.2	379 8	047959	CYTCHROME B.	2.08e+02
426	6	1.2	351 14	082485	HYPOHETICAL 39.3 KD P	2.08e+02	499	6	1.2	379 8	047959	CYTCHROME B.	2.08e+02
427	6	1.2	351 14	082485	NORTH AMERICAN SWINE I	2.08e+02	500	6	1.2	379 8	054623	STIMILAR TO SULOLOBUS S	2.08e+02
428	6	1.2	351 14	082485	HEMAGGLUTININ (FRAGMEN	2.08e+02	501	6	1.2	380 8	047976	CYTCHROME B.	2.08e+02
429	6	1.2	351 14	012677	HEMAGGLUTININ (FRAGMEN	2.08e+02	502	6	1.2	380 8	047961	CYTCHROME B.	2.08e+02
430	6	1.2	352 5	019220	F08G5.1 PROTEIN.	2.08e+02	503	6	1.2	382 8	021998	GALACTOYLASE.	2.08e+02
431	6	1.2	352 10	022443	SEED COAT PEROXIDASE P	2.08e+02	504	6	1.2	382 8	034340	CYTCHROME B LIGHT STR	2.08e+02
432	6	1.2	352 10	022443	HAELL RESTRICTION ENZY	2.08e+02	505	6	1.2	382 8	035538	CYTCHROME B LIGHT STR	2.08e+02
433	6	1.2	353 5	021051	COSMID F59G1.	2.08e+02	506	6	1.2	382 8	088515	CTCHROME B LIGHT STR	2.08e+02
434	6	1.2	353 14	067223	HA PRECURSOR (FRAGMENT	2.08e+02	507	6	1.2	382 8	088515	CTCHROME B LIGHT STR	2.08e+02
435	6	1.2	354 13	087499	MEL-1C(A) MELANONIN RE	2.08e+02	508	6	1.2	383 14	067012	HEMAGGLUTININ (FRAGMEN	2.08e+02
436	6	1.2	354 13	028281	IRON-SULFUR BINDING RE	2.08e+02	509	6	1.2	383 14	045938	HEMAGGLUTININ (FRAGMEN	2.08e+02
437	6	1.2	354 14	067735	INFLUENZA A/SL/2/87, A	2.08e+02	510	6	1.2	384 14	043055	HYPOHETICAL 42.3 KD P	2.08e+02
438	6	1.2	354 14	087496	MEL-1C(B) MELANONIN RE	2.08e+02	511	6	1.2	384 14	043055	HYPOHETICAL 42.3 KD P	2.08e+02
439	6	1.2	355 1	029857	CONSERVED HYPOHETICAL	2.08e+02	512	6	1.2	384 14	066931	GI GENE PRECURSOR.	2.08e+02
440	6	1.2	356 2	055216	CARMINOMYCIN 4-O-METHY	2.08e+02	513	6	1.2	385 14	067005	POLYPROTEIN PRECURSO	2.08e+02
441	6	1.2	356 2	028044	HYPOHETICAL 40.9 KD P	2.08e+02	514	6	1.2	385 14	009641	HEMAGGLUTININ PRECURSO	2.08e+02
442	6	1.2	357 6	018872	KERATINOCYTE GROWTH FA	2.08e+02	515	6	1.2	387 2	095416	HEMAGGLUTININ PRECURSO	2.08e+02
443	6	1.2	357 14	067091	HEMAGGLUTININ PRECURSO	2.08e+02	516	6	1.2	388 2	088027	NIJ.	2.08e+02
444	6	1.2	357 14	056948	PESTICIN.	2.08e+02	517	6	1.2	388 14	065360	HYPOHETICAL 41.6 KD P	2.08e+02
445	6	1.2	357 14	067090	HEMAGGLUTININ PRECURSO	2.08e+02	518	6	1.2	390 5	017274	HYPOHETICAL 42.1 KD P	2.08e+02
446	6	1.2	357 14	057159	PESTICIN.	2.08e+02	519	6	1.2	390 5	044148	12A1.4 PROTEIN.	2.08e+02
447	6	1.2	358 2	069073	HYPOHETICAL 37.8 KD P	2.08e+02	520	6	1.2	391 14	089925	C49A9.7 PROTEIN.	2.08e+02
448	6	1.2	358 2	069073	TRANSPOSASE.	2.08e+02	521	6	1.2	391 14	039170	PROBABLY SITE-SPECIFIC	2.08e+02
449	6	1.2	358 14	067088	HEMAGGLUTININ PRECURSO	2.08e+02	522	6	1.2	392 14	039173	HEMAGGLUTININ (FRAGME	2.08e+02
450	6	1.2	358 2	056916	TRSD.	2.08e+02	523	6	1.2	392 14	039168	HEMAGGLUTININ (FRAGME	2.08e+02
451	6	1.2	358 2	069960	POTATIVE TWO-COMPONENT	2.08e+02	524	6	1.2	393 5	045305	C47A10.3 PROTEIN.	2.08e+02
452	6	1.2	359 11	035476	GREEN OPSIN (GREEN-SEN	2.08e+02	525	6	1.2	393 5	053405	HYPOHETICAL 40.7 KD P	2.08e+02
453	6	1.2	360 4	093066	ENVELOPE PROTEIN (FRAG	2.08e+02	526	6	1.2	393 5	053405	SEPTINE/THEONINE KINAS	2.08e+02
454	6	1.2	360 4	093066	ENVELOPE PROTEIN (FRAG	2.08e+02	527	6	1.2	395 3	002340	ACTOXYDROXY-ACID ISOM	2.08e+02
455	6	1.2	360 4	093066	ENVELOPE PROTEIN (FRAG	2.08e+02	528	6	1.2	395 3	002340	ACTOXYDROXY-ACID ISOM	2.08e+02
456	6	1.2	360 4	093066	ENVELOPE PROTEIN (FRAG	2.08e+02	529	6	1.2	396 2	055794	HYPOHETICAL 44.6 KD P	2.08e+02
457	6	1.2	361 7	046882	MHC CLASS I DLA-88.	2.08e+02	530	6	1.2	397 2	074590	HYPOHETICAL 43.5 KD P	2.08e+02
458	6	1.2	361 7	046882	MHC CLASS I DLA-88.	2.08e+02	531	6	1.2	400 2	068009	HYPOHETICAL 40.7 KD P	2.08e+02

240	6	1.2	232	2	084263	DNA POL. III EPSILON CH	2.08e+02	313	2	034859	YKOV PROTEIN.	2.08e+02
241	6	1.2	233	9	064842	P. PROTEIN.	2.08e+02	314	2	033744	HYPOTHETICAL 33.8 KD P	2.08e+02
242	6	1.2	233	2	032558	ORF 1. PROTEIN.	2.08e+02	315	6	080500	F16B22.9 PROTEIN.	2.08e+02
243	6	1.2	233	2	P71005	HYPOTHETICAL 26.3 KD P	2.08e+02	316	6	088310	INFECTIN.	2.08e+02
244	6	1.2	236	2	P76942	PUTATIVE PROPHAGE SF6-	2.08e+02	317	6	014875	RNA BINDING PROTEIN.	2.08e+02
245	6	1.2	237	10	023436	SIMILARITY TO HYPOTHET	2.08e+02	318	4	041875	HOMEOBOX-LEUCINE ZIPPE	2.08e+02
246	6	1.2	243	11	088205	TESTIS SPECIFIC PROTEI	2.08e+02	319	10	040780	KILB PROTEIN.	2.08e+02
247	6	1.2	243	1	083006	PHOTOSYNTHETIC REACTIO	2.08e+02	320	6	057316	PROTEIN INVOLVED WITH	2.08e+02
248	6	1.2	244	2	005934	POTATIVE REGULATORY PR	2.08e+02	321	2	058165	320A LONG HYPOTHETICA	2.08e+02
249	6	1.2	245	2	000347	CL-9.	2.08e+02	322	6	046560	PYRIDOXAL KINASE (EC 2	2.08e+02
250	6	1.2	245	4	054045	POTATIVE CHAPERONE.	2.08e+02	323	5	018035	SIMILAR TO YEAST MAK16	2.08e+02
251	6	1.2	246	3	074648	ISOAMYL ACETATE HYDROL	2.08e+02	324	10	028153	TRANSCRIPTIONAL REGULA	2.08e+02
252	6	1.2	247	1	029253	HYPOTHETICAL 28.2 KD P	2.08e+02	325	5	041097	VEGETATIVE STORAGE PRO	2.08e+02
253	6	1.2	248	2	005437	HYPOTHETICAL 27.1 KD P	2.08e+02	326	5	019060	COSMID E04F6.	2.08e+02
254	6	1.2	249	2	P74442	HYPOTHETICAL 27.3 KD P	2.08e+02	327	14	083964	HAL CHAIN.	2.08e+02
255	6	1.2	249	9	080074	ANTI REPRESSOR.	2.08e+02	328	14	067115	HEMAGGLUTININ (FRAGEN	2.08e+02
256	6	1.2	251	9	034449	YOOD PROTEIN.	2.08e+02	329	14	045738	HEMAGGLUTININ (FRAGEN	2.08e+02
257	6	1.2	251	9	064113	POTATIVE ANTIREPRESSOR	2.08e+02	330	14	082490	T0365.5 PROTEIN.	2.08e+02
258	6	1.2	251	11	054895	GUANYLATE KINASE MEMBR	2.08e+02	331	14	009645	HEMAGGLUTININ (FRAGME	2.08e+02
259	6	1.2	252	11	026647	CONSERVED PROTEIN.	2.08e+02	332	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
260	6	1.2	252	5	016726	T07D3.2 PROTEIN.	2.08e+02	333	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
261	6	1.2	253	2	05668	HYPOTHETICAL 27.3 KD P	2.08e+02	334	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
262	6	1.2	253	2	020665	SIMILAR TO PHOSPHOLIPA	2.08e+02	335	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
263	6	1.2	256	5	022208	SIMILARITY TO GLYCOPRO	2.08e+02	336	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
264	6	1.2	257	11	089094	CASPASE-14.	2.08e+02	337	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
265	6	1.2	258	4	014400	GLUTAMATE DEHYDROGENAS	2.08e+02	338	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
266	6	1.2	259	5	062221	F33H2.8 PROTEIN.	2.08e+02	339	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
267	6	1.2	261	2	P72737	HYPOTHETICAL 29.5 KD P	2.08e+02	340	14	009649	HEMAGGLUTININ (FRAGMEN	2.08e+02
268	6	1.2	262	11	084648	TLS-ASSOCIATED PROTEIN	2.08e+02	341	14	067131	HEMAGGLUTININ (FRAGMEN	2.08e+02
269	6	1.2	262	4	075494	TLS-ASSOCIATED PROTEIN	2.08e+02	342	14	009643	HEMAGGLUTININ (FRAGMEN	2.08e+02
270	6	1.2	263	5	093783	F59F4.4 PROTEIN.	2.08e+02	343	14	009643	HEMAGGLUTININ (FRAGMEN	2.08e+02
271	6	1.2	263	13	090470	HOXD11.	2.08e+02	344	14	009643	HEMAGGLUTININ (FRAGMEN	2.08e+02
272	6	1.2	267	5	018233	SIMILARITY WITH WILMS.	2.08e+02	345	14	009643	HEMAGGLUTININ (FRAGMEN	2.08e+02
273	6	1.2	267	3	001138	CHITIN SYNTHASE (FRAGM	2.08e+02	346	14	027009	TUNGSTEN FORMYLMETANO	2.08e+02
274	6	1.2	268	2	083851	ABC TRANSPORTER, ATP-B	2.08e+02	347	14	039837	HEMAGGLUTININ (FRAGMEN	2.08e+02
275	6	1.2	269	1	029751	MOLYBDOPYRIN OXIDORED	2.08e+02	348	14	009643	HEMAGGLUTININ (FRAGMEN	2.08e+02
276	6	1.2	269	3	074305	POTATIVE SPLICING PROT	2.08e+02	349	14	009643	HEMAGGLUTININ (FRAGMEN	2.08e+02
277	6	1.2	278	2	053864	HYPOTHETICAL 30.1 KD P	2.08e+02	350	14	011317	HEMAGGLUTININ (FRAGMEN	2.08e+02
278	6	1.2	280	2	068106	COBALT TRANSPORT ATP-B	2.08e+02	351	14	044340	L-LACTATE DEHYDROGENAS	2.08e+02
279	6	1.2	281	5	023120	WO2B12.2 PROTEIN.	2.08e+02	352	14	066664	CASID PROTEIN.	2.08e+02
280	6	1.2	281	5	083580	HYPOTHETICAL 31.0 KD P	2.08e+02	353	14	066664	CASID PROTEIN.	2.08e+02
281	6	1.2	282	5	016874	C13A2.4 PROTEIN.	2.08e+02	354	14	048759	CUPC ATPASE (MEC).	2.08e+02
282	6	1.2	283	13	093417	CASPASE-3.	2.08e+02	355	14	014622	KILLER CELL RECEPTOR.	2.08e+02
283	6	1.2	283	10	042784	SEED COAT PEROXIDASE 1	2.08e+02	356	14	051379	HYPOTHETICAL 38.8 KD P	2.08e+02
284	6	1.2	284	6	099331	NEUTROPHIL PROTEIN (FR	2.08e+02	357	14	043803	NK RECEPTOR.	2.08e+02
285	6	1.2	284	6	007350	LIPASE ACTIVATOR PROTE	2.08e+02	358	14	082833	HEMAGGLUTININ (FRAGMEN	2.08e+02
286	6	1.2	284	11	062222	PAIRED BOX PROTEIN 6 (2.08e+02	359	14	067137	HEMAGGLUTININ (FRAGMEN	2.08e+02
287	6	1.2	290	2	P70999	HYPOTHETICAL 32.4 KD P	2.08e+02	360	14	084074	INFLUENZA A/SINAPORE/	2.08e+02
288	6	1.2	293	2	045076	HYDROXYMUTINOL 1,2-DIOX	2.08e+02	361	14	084074	INFLUENZA A/SINAPORE/	2.08e+02
289	6	1.2	293	10	038696	(CLONE KEXET-1) XET PR	2.08e+02	362	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
290	6	1.2	294	10	081290	T14P8.21 PROTEIN.	2.08e+02	363	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
291	6	1.2	295	14	086360	G PROTEIN (FRAGMENT).	2.08e+02	364	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
292	6	1.2	295	6	077621	ALPHA-1A ADRENORECEPT	2.08e+02	365	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
293	6	1.2	295	2	066147	HYPOTHETICAL 31.0 KD P	2.08e+02	366	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
294	6	1.2	297	14	091947	ATTACHMENT PROTEIN.	2.08e+02	367	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
295	6	1.2	297	14	082071	GLYCOPROTEIN.	2.08e+02	368	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
296	6	1.2	297	14	082071	GLYCOPROTEIN.	2.08e+02	369	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
297	6	1.2	297	14	082071	GLYCOPROTEIN.	2.08e+02	370	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
298	6	1.2	297	14	082067	GLYCOPROTEIN.	2.08e+02	371	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
299	6	1.2	297	14	082067	GLYCOPROTEIN.	2.08e+02	372	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
300	6	1.2	298	14	082068	GLYCOPROTEIN (FRAGMENT	2.08e+02	373	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
301	6	1.2	298	14	001929	MAJOR SURFACE GLYCOPRO	2.08e+02	374	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
302	6	1.2	298	14	058205	298A LONG HYPOTHETICA	2.08e+02	375	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
303	6	1.2	300	10	023734	SIMILAR TO NAM.	2.08e+02	376	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
304	6	1.2	301	9	021879	MAJOR STRUCTURAL PROTE	2.08e+02	377	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
305	6	1.2	301	9	080113	POTATIVE MAJOR STRUCTU	2.08e+02	378	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
306	6	1.2	303	2	P72971	HYPOTHETICAL 33.9 KD P	2.08e+02	379	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
307	6	1.2	303	2	066882	TRANSCRIPTIONAL REGULA	2.08e+02	380	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
308	6	1.2	306	10	045286	ORF 2.	2.08e+02	381	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
309	6	1.2	309	1	023477	HYPOTHETICAL 33.5 KD P	2.08e+02	382	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
310	6	1.2	310	5	001301	CONSERVED HYPOTHETICAL	2.08e+02	383	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
311	6	1.2	310	10	065269	F25H9.2 PROTEIN.	2.08e+02	384	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
312	6	1.2	311	2	068976	F6N23.28 PROTEIN.	2.08e+02	385	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
313	6	1.2	311	2	068976	HYPOTHETICAL 35.1 KD P	2.08e+02	386	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02

94	7	1.4	869	5	001262	F49C12.15 PROTEIN.	3.55e+00	167	6	1.2	133	10	041863	TRANSPSABLE ELEMENT M	2.08e+02
95	7	1.4	875	5	018372	C33D9.1 PROTEIN.	3.55e+00	168	6	1.2	135	3	013568	YPL130CP.	2.08e+02
96	7	1.4	884	4	000302	ERRPOT 213-21.	3.55e+00	169	6	1.2	141	2	084321	L11 RIBOSOMAL PROTEIN.	2.08e+02
97	7	1.4	920	4	014673	KIA0164 PROTEIN.	3.55e+00	170	6	1.2	144	2	025757	ATP SYNTHASE F0, SUBUN	2.08e+02
98	7	1.4	932	5	017753	SIMILAR TO KINESIN-LIK	3.55e+00	171	6	1.2	145	9	038222	ORF.	2.08e+02
99	7	1.4	952	5	094889	PROTEIN KINASE.	3.55e+00	172	6	1.2	145	5	038222	COSMID C05H8.	2.08e+02
100	7	1.4	967	3	074561	PUTATIVE B-ZIP TRANSCR	3.55e+00	173	6	1.2	148	2	083123	PPTS SYNTHASE, NITROGEN	2.08e+02
101	7	1.4	1032	11	062760	RNA HELICASE.	3.55e+00	174	6	1.2	151	14	010614	ACNPV ORF1 HOMOLOG.	2.08e+02
102	7	1.4	1044	10	080988	PUTATIVE GLYCINE DEHYD	3.55e+00	175	6	1.2	153	2	032455	ORF 5.	2.08e+02
103	7	1.4	1048	11	063627	CTD-BINDING SR-LIKE PR	3.55e+00	176	6	1.2	153	2	024931	CONSERVED HYPOHETICAL	2.08e+02
104	7	1.4	1072	5	045322	COLCT.1 PROTEIN.	3.55e+00	177	6	1.2	155	5	021739	R05D11.7 PROTEIN.	2.08e+02
105	7	1.4	1125	3	007660	CHROMOSOME IV READING	3.55e+00	178	6	1.2	161	10	P93490	CELL WALL INVERTASE II	2.08e+02
106	7	1.4	1150	3	002511	HYPOHETICAL 133.0 KD	3.55e+00	179	6	1.2	163	5	017155	ALPHA CHAIN OF THE TET	2.08e+02
107	7	1.4	1151	13	P87343	MEMBRANE GUANYLYL CYCL	3.55e+00	180	6	1.2	163	14	072789	ENVELOPE PROTEIN (FRAG	2.08e+02
108	7	1.4	1173	14	098177	MC0061.	3.55e+00	181	6	1.2	163	15	094442	TENTANS ORF'S (A-E) FO	2.08e+02
109	7	1.4	1173	11	063624	CTD-BINDING SR-LIKE PR	3.55e+00	182	6	1.2	163	14	072791	ENVELOPE PROTEIN (FRAG	2.08e+02
110	7	1.4	1175	14	085285	ORF17.	3.55e+00	183	6	1.2	163	4	025410	NUCLEIC ACID BINDING P	2.08e+02
111	7	1.4	1179	4	014120	DBP-5 NUCLEAR PROTEIN.	3.55e+00	184	6	1.2	163	14	072772	ENVELOPE PROTEIN (FRAG	2.08e+02
112	7	1.4	1246	3	012276	CHROMOSOME XV READING	3.55e+00	185	6	1.2	163	14	072788	ENVELOPE PROTEIN (FRAG	2.08e+02
113	7	1.4	1257	10	064768	PUTATIVE SER/THR PROTE	3.55e+00	186	6	1.2	163	14	072790	ENVELOPE PROTEIN (FRAG	2.08e+02
114	7	1.4	1268	11	063623	CTD-BINDING SR-LIKE PR	3.55e+00	187	6	1.2	163	14	072773	ENVELOPE PROTEIN (FRAG	2.08e+02
115	7	1.4	1280	4	075158	KIA0670 PROTEIN (FRAG	3.55e+00	188	6	1.2	166	8	079557	NADH DEHYDROGENASE SUB	2.08e+02
116	7	1.4	1288	4	015038	KIA00524 (FRAGMENT).	3.55e+00	189	6	1.2	171	2	053838	HYPOHETICAL PROTEIN (2.08e+02
117	7	1.4	1308	4	060307	KIA00561 PROTEIN (FRAG	3.55e+00	190	6	1.2	172	14	P69177	GLYCOPROTEIN M (GM) (F	2.08e+02
118	7	1.4	1471	5	018245	C27B7.7 PROTEIN.	3.55e+00	191	6	1.2	172	14	P87513	POLYMERASE NSP4 (FRAGM	2.08e+02
119	7	1.4	1590	5	P91152	SIMILAR TO DROSOPHILA	3.55e+00	192	6	1.2	173	1	026725	HYPOHETICAL 19.6 KD P	2.08e+02
120	7	1.4	1791	4	060382	KIA0324 (FRAGMENT).	3.55e+00	193	6	1.2	176	4	025380	Y-CHROMOSOME RNA RECOG	2.08e+02
121	7	1.4	1926	11	061464	NUCLEAR PROTEIN, NP220	3.55e+00	194	6	1.2	178	2	031285	PUTATIVE INNERMEMBRANE	2.08e+02
122	7	1.4	1935	3	060072	RNA HELICASE.	3.55e+00	195	6	1.2	178	5	062303	K11D2.1 PROTEIN.	2.08e+02
123	7	1.4	1978	4	014966	NUCLEAR PROTEIN, NP220	3.55e+00	196	6	1.2	179	14	067328	INFLUENZA A/PR/8/34, H	2.08e+02
124	7	1.4	2161	4	001668	NEURONAL L-TYPE CALCIO	3.55e+00	197	6	1.2	180	10	064541	YU8B12R.28.	2.08e+02
125	7	1.4	2170	4	013931	VOLTAGE-DEPENDENT CALC	3.55e+00	198	6	1.2	182	6	077625	CASPASE-3 (FRAGMENT).	2.08e+02
126	7	1.4	2181	4	013916	NEUROENDOCRINE/BETA-CE	3.55e+00	199	6	1.2	183	14	067650	TYPE III P20 PROTEIN.	2.08e+02
127	7	1.4	2783	5	P91255	FL2F3.2 PROTEIN.	3.55e+00	200	6	1.2	183	11	070307	TLS-ASSOCIATED PROTEIN	2.08e+02
128	7	1.4	4128	11	088187	DNA-DEPENDENT PROTEIN	3.55e+00	201	6	1.2	183	4	060572	TL5-ASSOCIATED PROTEIN	2.08e+02
129	7	1.4	4128	11	P97313	PAIRED BOX PROTEIN PAX	3.55e+00	202	6	1.2	183	5	005817	HYPOHETICAL 19.5 KD P	2.08e+02
130	6	1.2	39	4	007378	RETINOBLASTOMA SUSCEPT	2.08e+02	203	6	1.2	184	5	016158	CALCIUM-BINDING PROTEI	2.08e+02
131	6	1.2	41	4	092700	RETINOBLASTOMA SUSCEPT	2.08e+02	204	6	1.2	184	11	P97297	PREFRONDENOMEDULLIN P	2.08e+02
132	6	1.2	42	14	010425	HEMAGGLUTININ (FRAGMEN	2.08e+02	205	6	1.2	184	11	P97453	ADRENOMEDULLIN.	2.08e+02
133	6	1.2	44	8	032935	ORF44.	2.08e+02	206	6	1.2	186	2	P73747	HYPOHETICAL 21.2 KD P	2.08e+02
134	6	1.2	44	8	032935	PROTAMINE 2.	2.08e+02	207	6	1.2	188	2	P87042	CLASS I CHITIN SYNTHAS	2.08e+02
135	6	1.2	49	13	013031	ATTACHMENT GLYCOPROTEI	2.08e+02	208	6	1.2	188	3	007361	HESA.	2.08e+02
136	6	1.2	52	14	092312	(CLONE E35) GENE FROM	2.08e+02	209	6	1.2	189	4	P87044	CLASS II CHITIN SYNTHA	2.08e+02
137	6	1.2	70	4	014070	ORE73B.	2.08e+02	210	6	1.2	189	4	007823	MAG30 PROTEIN (FRAGMEN	2.08e+02
138	6	1.2	73	8	033001	GRPASE A (FRAGMENT).	2.08e+02	211	6	1.2	190	2	052136	ESCG.	2.08e+02
139	6	1.2	73	8	P71322	GRPASE A (FRAGMENT).	2.08e+02	212	6	1.2	195	8	034670	CYTOSOLIC B (FRAGMENT	2.08e+02
140	6	1.2	77	1	026574	CONSERVED PROTEIN.	2.08e+02	213	6	1.2	196	2	086675	HYPOHETICAL 21.4 KD P	2.08e+02
141	6	1.2	78	2	069048	RYRMB2.	2.08e+02	214	6	1.2	197	11	088375	PUTATIVE KERATIN-ASSOC	2.08e+02
142	6	1.2	80	2	069048	GRPASE A (FRAGMENT).	2.08e+02	215	6	1.2	199	14	067675	VIRUS COAT PROTEIN.	2.08e+02
143	6	1.2	81	1	054608	SIMILAR TO BACILLUS SU	2.08e+02	216	6	1.2	199	14	065970	COAT PROTEIN.	2.08e+02
144	6	1.2	87	2	032324	PUTATIVE CELL DIVISION	2.08e+02	217	6	1.2	207	5	076587	FL6G10.5 PROTEIN.	2.08e+02
145	6	1.2	90	5	016396	D2062.6 PROTEIN.	2.08e+02	218	6	1.2	212	8	003357	ATP SYNTHASE A CHAIN (2.08e+02
146	6	1.2	95	14	005535	NEUTROPHIL PROTEIN (FR	2.08e+02	219	6	1.2	213	8	035425	CYTOSOLIC B (FRAGMENT	2.08e+02
147	6	1.2	99	4	015789	H+-TRANSPORTING ATP SY	2.08e+02	220	6	1.2	213	14	077629	ENVELOPE GLYCOPROTEIN	2.08e+02
148	6	1.2	101	10	023452	SIMILARITY TO RIBOSOMA	2.08e+02	221	6	1.2	214	14	011784	VIF PROTEIN.	2.08e+02
149	6	1.2	101	10	023452	HYPOHETICAL 13.0 KD P	2.08e+02	222	6	1.2	214	14	086926	TRANSMEMBRANE PROTEIN	2.08e+02
150	6	1.2	107	8	078941	NITROGENASE (FRAGMENT)	2.08e+02	223	6	1.2	214	3	043039	HYPOHETICAL 24.7 KD P	2.08e+02
151	6	1.2	108	2	054032	(CJS193) MITOCHONDRIAL	2.08e+02	224	6	1.2	216	3	053857	INTERNAL ORF TO 05657.	2.08e+02
152	6	1.2	114	8	033757	(CJS566) MITOCHONDRIAL	2.08e+02	225	6	1.2	216	6	053856	PUTATIVE RESPONSE REGU	2.08e+02
153	6	1.2	115	2	056549	ALPHA SUBUNIT OF DINIT	2.08e+02	226	6	1.2	216	4	043884	OLFACTORY RECEPTOR (FR	2.08e+02
154	6	1.2	117	10	040816	S-RNASE (FRAGMENT).	2.08e+02	227	6	1.2	216	14	089753	EHO. COMPLETE GENOME.	2.08e+02
155	6	1.2	117	10	040816	S-RNASE (FRAGMENT).	2.08e+02	228	6	1.2	217	2	051108	TRANSFERIN BINDING PR	2.08e+02
156	6	1.2	117	10	066038	HOST SHUT OFF VIRION P	2.08e+02	229	6	1.2	220	4	015376	Y-CHROMOSOME RNA RECOG	2.08e+02
157	6	1.2	119	14	071360	ENVELOPE GLYCOPROTEIN,	2.08e+02	230	6	1.2	220	4	051814	HYPOHETICAL 25.1 KD P	2.08e+02
158	6	1.2	121	6	P79109	G-PROTEIN-COUPLED INWA	2.08e+02	231	6	1.2	223	5	094611	CYCLOPHILIN.	2.08e+02
159	6	1.2	121	5	015618	ELONGATION FACTOR 1 BE	2.08e+02	232	6	1.2	223	11	026310	DNA-DEPENDENT DNA POLY	2.08e+02
160	6	1.2	123	10	040821	CYTOSOLIC B (FRAGMENT	2.08e+02	233	6	1.2	224	11	008440	TRANSCRIPTION FACTOR E	2.08e+02
161	6	1.2	124	8	078920	CYTOSOLIC B (FRAGMENT	2.08e+02	234	6	1.2	225	5	061815	B0511.5 PROTEIN.	2.08e+02
162	6	1.2	124	6	095294	NUCLEAR PROTEIN (FRAGM	2.08e+02	235	6	1.2	225	4	043877	OLFACTORY RECEPTOR (FR	2.08e+02
163	6	1.2	124	6	046582	CYTOSOLIC C OXIDASE P	2.08e+02	236	6	1.2	227	14	011286	HYPOHETICAL 24.6 KD P	2.08e+02
164	6	1.2	125	13	057510	OOCTE MATURATION FACT	2.08e+02	237	6	1.2	230	4	099561	NK RECEPTOR.	2.08e+02
165	6	1.2	125	5	005757	HOMEBOX PROTEIN ABDOM	2.08e+02	238	6	1.2	231	4	014911	TAX INTERACTION PROTEI	2.08e+02
166	6	1.2	125	13	057511	OOCTE MATURATION FACT	2.08e+02	239	6	1.2	231	2	069442	SUCRASE (EC 3.2.1.26)	2.08e+02


```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9436758.
RA WONG W.T., CARLOMAGNO F., DRUCK T., BARLETTA C., CROCE C.M.,
RA HUBNER K., KRAUS M.H., DI FIORE P.P.;
RT "Evolutionary conservation of the EPS8 gene and its mapping to human
RT chromosome 12q23-q24."
RL ONCOGENE 9:3057-3061(1994).
CC -1- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT
CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.
CC -1- PPM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES ANALYZED, INCLUDING
CC HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND
CC PANCREAS. EXPRESSED IN ALL EPITHELIAL AND FIBROBLASTIC LINES
CC EXAMINED AND IN SOME, BUT NOT ALL, HEMATOPOIETIC CELLS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SPLIT PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12535; G530823; -.
DR MIM: 600206; -.
DR PROSITE: PS50002; SH3; 1.
DR PRAM: PF00018; SH3; 1.
DR HSSP: Q08509; 1A0J.
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 69 129 PH (FIRST PART).
FT DOMAIN 210 213 POLY-PRO.
FT DOMAIN 322 325 POLY-PRO.
FT DOMAIN 381 414 PH (SECOND PART).
FT DOMAIN 421 440 PRO-RICH.
FT DOMAIN 532 591 SH3.
FT DOMAIN 615 651 PRO-RICH.
FT DOMAIN 659 664 POLY-SER.
SV SEQUENCE 822 AA; 91881 MW; 501CF5AE CRC32;

Query Match 1.6%; Score 8; DB 1; Length 822;
Best Local Similarity 100.0%; Pred. No. 2.00e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

53 RDSVSVS 60
11111111
QY 130 RDSVSVS 137

RLSULI 15
ID EMI_ENSMI STANDARD; PRT; 34 AA.
AC P27205;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE SPERM PROTEIN EMI (FRAGMENT).
OS ENSIS MINOR (RAZOR SHELL) (MINOR JACKKNIFE CLAM).
OC EUKARYOTA; METAZOA; MOLUSCA; BIVALVIA; HETEROCONCHIA; VENEROIDA;
OC SOLENIIDAE; ENSIS.
RN [1]
RP SEQUENCE.
RC TISSUE-SPERM;
RX MEDLINE: 92190249.
RT "Molluscan sperm proteins: Ensis minor."
RL BIOCHIM. BIOPHYS. ACTA 1119:296-302(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: PROLAMINE-LIKE.
DR PIR: S21080; S21080.
KW CHROMOSOMAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; SPERM; REPEAT.

```

```

FT DOMAIN 3 16 7 X 2 AA TANDEM REPEATS OF S-[KR].
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3531 MW; 77A8D51E CRC32;

Query Match 1.4%; Score 7; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 2.20e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 RSRSR 12
1111111
QY 150 RSRSR 156

Search completed: Tue Aug 10 11:39:50 1999
Job time : 77 secs.

```

DR EMBL: D90913; G1653440; -
 DR PROSITE: PS00933; FGGY_KINASES_1; FALSE_NEG.
 DR PROSITE: PS00445; FGGY_KINASES_2; 1.
 DR PFAM: PF00370; FGGY; 1.
 DR HSSP: P08859; IGLB.
 KW GLYCEROL METABOLISM; TRANSFERASE; KINASE.
 SQ SEQUENCE 495 AA; 54351 MW; C7247592 CRC32;

Query Match 1.6%; Score 8; DB 1; Length 495;
 Best Local Similarity 100.0%; Pred. No. 2,00e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 338 LARSVDN 345
 11111111
 QY 91 LARSVDN 98

RESULT 12
 ID VE2_HP47 STANDARD; PRT; 506 AA.
 AC P22420;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE REGULATORY PROTEIN E2.
 GN E2.
 OS HUMAN PAPILLOMAVIRUS TYPE 47.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPAPOVIRIDAE; PAPILLOMAVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90281611.
 KW Kiyono T., Adachi A., Ishibashi M.;
 RT "Genome organization and taxonomic position of human papillomavirus
 type 47 inferred from its DNA sequence.";
 RL VIROLOGY 177:401-405(1990).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2R RESPONSE ELEMENT (5'-ACCCNNNNNGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2R'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M33305; G333067; -
 DR PIR: D35324; W2M47.
 DR PFAM: PF00508; E2_N; 1.
 DR PFAM: PF00511; E2_C; 1.
 DR HSSP: P03122; 2BOP.
 KW EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;
 KW TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN.
 SQ SEQUENCE 506 AA; 57478 MW; F93097EF CRC32;

Query Match 1.6%; Score 8; DB 1; Length 506;
 Best Local Similarity 100.0%; Pred. No. 2,00e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 302 RARSRS 309
 11111111
 QY 148 RARSRS 155

RESULT 13
 ID EPS8_MOUSE STANDARD; PRT; 821 AA.

AC 008509;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8.
 GN EPS8.
 OS MUS MUSCULUS (MUS).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94008987.
 RA FAZIOLE F., MINICHELLO L., MATOSKA V., CASTAGNINO P., MITI T.,
 RA WONG W.T., DI FIORE P.P.;
 RT "Eps8, a substrate for the epidermal growth factor receptor kinase,
 RT enhances EGF-dependent mitogenic signals.";
 RL EMBO J. 12:3799-3808(1993).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 532-591.
 RX MEDLINE: 97448677.
 RA KISHAN K.V.R., SCITA G., WONG W.T., DI FIORE P.P., NEWCOMER M.E.;
 RT "The SH3 domain of Eps8 exists as a novel intertwined dimer.";
 RL NAT. STRUCT. BIOL. 4:739-743(1997).
 CC -1- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT
 CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.
 CC -1- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L21671; G309217; -
 DR PDB: 1A0J; 08-JUL-98.
 DR MGD: MG1:104684; EPS8.
 DR PROSITE: PS50002; SH3; 1.
 DR PFAM: PF00018; SH3; 1.
 KW SH3 DOMAIN; PHOSPHORYLATION; 3D-STRUCTURE.
 FT DOMAIN 69 129
 FT DOMAIN 210 213
 FT DOMAIN 322 325
 FT DOMAIN 381 414
 FT DOMAIN 421 440
 FT DOMAIN 532 591
 FT DOMAIN 620 650
 FT DOMAIN 658 663
 SQ SEQUENCE 821 AA; 91738 MW; 56D28642 CRC32;

Query Match 1.6%; Score 8; DB 1; Length 821;
 Best Local Similarity 100.0%; Pred. No. 2,00e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 RDSVSSVS 60
 11111111
 QY 130 RDSVSSVS 137

RESULT 14
 ID EPS8_HUMAN STANDARD; PRT; 822 AA.
 AC Q12929;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8.
 GN EPS8.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

CC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 CC PLECOGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
 CC DROSOPHILIDAE; DROSOPHILA.
 CC (1)
 CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 CC TISSUE-EMBRYO;
 CC MEDLINE: 96413646.
 CC RUDNER D.Z., KANAR R., BRERER K.S., RIO D.C.;
 CC "Mutations in the small subunit of the Drosophila U2AF splicing
 CC factor cause lethality and developmental defects";
 CC PROC. NATL. ACAD. SCI. U.S.A. 93:10333-10337(1996).
 CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. BINDS TO THE
 CC POLYRIMIDINE TRACT OF INTRONS EARLY DURING SPLICOSOME ASSEMBLY
 CC (BY SIMILARITY).
 CC -1- SUBUNIT: ASSOCIATES WITH A 65 KD PROTEIN (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC CC EMBL: U67066; G1621615; -
 CC DR FLXBASE; FB990017457; U2af38.
 CC DR PFAM; PF00076; rrm_1.
 CC DR PFAM; PF00642; zt-ccch; 2.
 CC KW NUCLEAR PROTEIN; RNA-BINDING; MRNA SPLICING.
 CC FT DOMAIN 46 174 INVOLVED IN BINDING U2AF-50
 CC FT DOMAIN 180 213 (BY SIMILARITY).
 CC FT DOMAIN 190 197 ARG/SER-RICH (RS DOMAIN).
 CC FT DOMAIN 252 262 POLY-ARG.
 CC FT DOMAIN 252 262 POLY-GLY.
 CC SO SEQUENCE 264 AA; 29855 MW; C267A411 CRC32;
 CC -----
 CC Query Match 1.6%; Score 8; DB 1; Length 264;
 CC Best Local Similarity 100.0%; Pred. No. 2.00e-02;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC Db 193 RARSRS 200
 CC QY 148 RARSRS 155
 CC -----
 CC RESULT 10
 CC ID ICE3_XENLA STANDARD: PRT: 282 AA.
 CC AC P55866;
 CC DT 01-NOV-1997 (REL. 35, CREATED)
 CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 CC DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 CC DE APOPTAIN PRECURSOR (EC 3.4.22.-) (CYSTEINE PROTEASE CPP32) (CASPASE-3)
 CC (CASP-3) (XCP32).
 CC GN CASP3.
 CC OS XENOPUS LAEVIS (AFRICAN CLAMMED FROG)
 CC CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
 CC CC MESOBATRACHIA; PIPOIDEA; PIPOIDAE; XENOPODINAE; XENOPUS.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 97184166.
 CC RA YAOITA Y., NAKAJIMA K.;
 CC RT "Induction of apoptosis and CPP32 expression by thyroid hormone in a
 CC myoblastic cell line derived from tadpole tail";
 CC J. BIOL. CHEM. 273:5122-5127(1997).
 CC -1- FUNCTION: IMPORTANT MEDIATOR OF APOPTOSIS. AT THE ONSET OF
 CC APOPTOSIS IT PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE
 CC (PARP) AT A 216-ASP-1-GLY-217 BOND (P17) AND A 12 KD (P12) SUBUNITS.
 CC -1- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY)
 CC -----
 CC THE SUBUNITS ARE DERIVED FROM THE PRECURSOR SEQUENCE BY A PROBABLE

CC AUTOCATALYTIC MECHANISM AND PROBABLY BY OTHER CASPASES (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
 CC CASPASE FAMILY.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC CC EMBL: D89784; D1014718; -
 CC DR PROSITE; PS01121; CASPASE_HIS; 1.
 CC DR PROSITE; PS01122; CASPASE_CYS; 1.
 CC DR PFAM; PF00655; ICE_P10; 1.
 CC DR PFAM; PF00656; ICE_P20; 1.
 CC DR HSSP; P42574; IPAU.
 CC KW HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
 CC FT PROPEP 1 2
 CC FT CHAIN 186
 CC FT CHAIN 282
 CC FT ACT SITE 187 282 APOPTAIN P12 SUBUNIT.
 CC FT ACT SITE 131 131 BY SIMILARITY.
 CC FT ACT SITE 174 174 BY SIMILARITY.
 CC SO SEQUENCE 282 AA; 32124 MW; E9608701 CRC32;
 CC -----
 CC Query Match 1.6%; Score 8; DB 1; Length 282;
 CC Best Local Similarity 100.0%; Pred. No. 2.00e-02;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC Db 56 GMCLINN 63
 CC QY 243 GMCLINN 250
 CC -----
 CC RESULT 11
 CC ID GLPK_SYNY3 STANDARD: PRT: 495 AA.
 CC AC P74260;
 CC DT 01-NOV-1997 (REL. 35, CREATED)
 CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 CC DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)
 CC (GLYCEROKINASE) (GK).
 CC DE GLPK OR SLR1672.
 CC GN SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 CC OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 CC CC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 97061201.
 CC RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMITU E., NAKAMURA Y.,
 CC RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKI T., KIMURA T.,
 CC RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARDO K.,
 CC RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,
 CC RA YAMADA M., YASUDA M., TABATA S.;
 CC RT "Sequence analysis of the genome of the unicellular cyanobacterium
 CC Synechocystis sp. strain PCC6803. II. Sequence determination of the
 CC entire genome and assignment of potential protein-coding regions";
 CC J. DNA RES. 3:109-136(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.
 CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
 CC GLYCEROKINASE / XYLOKINASE FAMILY.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

RT "Purification, cloning, and expression of ciliary neurotrophic factor
 RT (CNTF).";
 RT SCIENCE 246:1023-1025(1989).
 RN [2]
 RX PARTIAL SEQUENCE.
 RA MEDLINE: 90256829.
 RA LIN L.-F.H., ARMES L.G., SOMMER A., SMITH D.J., COLLINS F.;
 RT "Isolation and characterization of ciliary neurotrophic factor from
 RT rabbit sciatic nerves";
 RL J. BIOL. CHEM. 265:8942-8947(1990).
 CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
 CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
 CC AXOTOMY.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M29828; G164890; -
 DR PIR: A40082; UNRBCF.
 DR PFAM: PF01110; CNTF. 1.
 DR HSSP: P26441; ICNT.
 KM GROWTH FACTOR; NEURONE.
 SO SEQUENCE 199 AA; 22662 MW; BB9A0976 CRC32;
 Query Match 1.6%; Score 8; DB 1; Length 199;
 Best Local Similarity 100.0%; Pred. No. 2,00e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 50 LDSVDGVP 57
 OY 370 LDSVDGVP 377
 ID CNTF_PIG STANDARD; PRT; 200 AA.
 AC 002732;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE CILIARY NEUROTROPHIC FACTOR (CNTF).
 GN CNTF.
 OS SUS SCROFA (PIG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLIA; SUIFORMES; SUINA; SUIDAE; SUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA PIEDRAHITA J.A., WEAKS R.;
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
 CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
 CC AXOTOMY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM (BY SIMILARITY).
 CC -1- SIMILARITY: HIGH, WITH OTHER MAMMALIAN CNTF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U57644; G2072395; -
 DR PFAM: PF01110; CNTF. 1.

DR HSSP: P26441; ICNT.
 KM GROWTH FACTOR; NEURONE.
 SO SEQUENCE 200 AA; 22718 MW; CDFEEB09 CRC32;
 Query Match 1.6%; Score 8; DB 1; Length 200;
 Best Local Similarity 100.0%; Pred. No. 2,00e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 50 LDSVDGVP 57
 OY 370 LDSVDGVP 377
 ID CNTF_RAT STANDARD; PRT; 200 AA.
 AC P20294;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE CILIARY NEUROTROPHIC FACTOR (CNTF).
 GN CNTF.
 OS RATUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 90081871.
 RA STOECKLI K.A., LOTTSPREICH F., SENDTNER M., MASTAKOWSKI P., CARROLL P.,
 RA GOETZ R., LINDHOLM D., THOENEN H.;
 RT "Molecular cloning, expression and regional distribution of rat
 RT ciliary neurotrophic factor";
 RL NATURE 342:920-923(1989).
 CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
 CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
 CC AXOTOMY.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X17457; G55969; -
 DR PIR: S08144; UNRBCF.
 DR PFAM: PF01110; CNTF. 1.
 DR HSSP: P26441; ICNT.
 KM GROWTH FACTOR; NEURONE.
 SO SEQUENCE 200 AA; 22854 MW; A1693EP9 CRC32;
 Query Match 1.6%; Score 8; DB 1; Length 200;
 Best Local Similarity 100.0%; Pred. No. 2,00e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 50 LDSVDGVP 57
 OY 370 LDSVDGVP 377
 ID U2AG_DROME STANDARD; PRT; 264 AA.
 AC 094535;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE SPLICING FACTOR U2AF 38 KD SUBUNIT (U2 AUXILIARY FACTOR 38 KD
 DE SUBUNIT) (U2 SNRNP AUXILIARY FACTOR SMALL SUBUNIT).
 GN U2AF38.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).

OS CAENORHABDITIS VULGARIS.
 OC EUKARYOTA: METAZOA: NEMATODA: SECRETENTERA: RHABDITIA, RHABDITIDA;
 OC RHABDITINA: RHABDITOIDEA: RHABDITIDAE: PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94061982.
 RA YUAN J., SHAHAM S., LEDOUX S., ELLIS H.M., HORVITZ H.R.;
 RT "The C. elegans cell death gene ced-3 encodes a protein similar to
 RT mammalian interleukin-1 beta-converting enzyme.";
 RL CELL 75:641-652(1993).
 CC -1- FUNCTION: ACTS AS A CYSTEINE PROTEASE IN CONTROLLING PROGRAMMED
 CC CELL DEATH BY PROTEOLYTICALLY ACTIVATING OR INACTIVATING A
 CC SUBSTRATE PROTEIN OR PROTEINS. A POTENTIAL SUBSTRATE MAY BE CED-4.
 CC ALTERNATIVELY IT MIGHT DIRECTLY CAUSE CELL DEATH BY
 CC PROTEOLYTICALLY CLEAVING PROTEINS THAT ARE CRUCIAL FOR CELL
 CC VIABILITY (BY SIMILARITY).
 CC -1- SUBUNIT: COULD BE A HETERODIMER OF TWO SUBUNITS DERIVED FROM THE
 CC PRECURSOR SEQUENCE BY A PROBABLE AUTOCATALYTIC MECHANISM.
 CC -1- PTM: MAY BE REGULATED BY PHOSPHORYLATION.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
 CC CASPASE FAMILY.
 CC PROSITE: PS01121; CASPASE_HIS: 1.
 CC PROSITE: PS01122; CASPASE_CYS: 1.
 CC DR PFAM: PF00619; CARD: 1.
 CC DR PFAM: PF00655; ICE_P10: 1.
 CC DR PFAM: PF00656; ICE_P20: 1.
 CC DR HSSP: P42574; ICP3.
 CC HYDROLASE: THIOLESTERASE; ZINCEN; APOPTOSIS; PHOSPHORYLATION.
 CC FT CHAIN 1 364
 CC FT CHAIN 1 364
 CC FT CHAIN 365 496
 CC FT ACT_SITE 308 308
 CC FT ACT_SITE 351 351
 CC FT SEQUENCE 496 AA; 55945 MM; CBDF84D CRC32:
 CC Query Match 14.3%; Score 72; DB 1; Length 496;
 CC Best Local Similarity 100.0%; Pred. No. 2.05e-222;
 CC Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 394 QOVRKRPQSDILIRATTAQVYVSNRSAGSFIVANCEVSTAKKMDVYELLTEVN 453
 QY 403 QOVRKRPQSDILIRATTAQVYVSNRSAGSFIVANCEVSTAKKMDVYELLTEVN 462
 DB 454 KKVACGFOTSOG 465
 QY 463 KKVACGFOTSOG 474

RESULT 3
 ID ICE7_HUMAN STANDARD; PRT; 303 AA.
 AC P55210; Q13364;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE CASPASE-7 PRECURSOR (EC 3.4.22.-) (ICE-LIKE APOPTOTIC PROTEASE 3)
 GN (ICE-LAB3) (APOPTOTIC PROTEASE MCH-3) (CMH-1).
 OS CASP7 OR MCH3.
 OS HOMO SAPIENS (HUMAN).
 OS EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA;
 OC PRIMATES: CATARRHINI: HOMINIDAE: HOMO.
 RN [1]
 RP SEQUENCE FROM N.A. (ALPHA ISOFORM).
 RX MEDLINE: 96139498.
 RA DUAN H., CHINNAIYAN A.M., HUDSON P.L., WING J.P., HE W.-W.,
 RA DIXIT V.M.;
 RT "ICE-LAB3, a novel mammalian homologue of the Caenorhabditis elegans
 RT cell death protein Ced-3 is activated during Fas- and tumor necrosis
 RT factor-induced apoptosis.";
 RL J. BIOL. CHEM. 271:1621-1625(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ALPHA ISOFORM).
 RX TISSUE-SPLEEN;

RX MEDLINE: 96147144.
 RA LIPKKE J.A., GU Y., SARNECKI C., CARON P.R., SU M.S.-S.;
 RT "Identification and characterization of CPP32/Mch2 homolog 1, a novel
 RT cysteine protease similar to CPP32.";
 RL J. BIOL. CHEM. 271:1825-1828(1996).
 RN [3]
 RP SEQUENCE FROM N.A. (ALPHA AND BETA ISOFORMS).
 RC TISSUE-T CELL;
 RX MEDLINE: 9615019.
 RA FERNANDES-ALNEMRI T., TAKAHASHI A., ARMSTRONG R.C., KREBS J.,
 RA FRITZ L., TOMASELLI K.J., WANG J., YU Z., CROCE C.M., SALVESEN G.,
 RA EARNshaw W.C., LITWACK G., ALNEMRI E.S.;
 RT "Mch3, a novel human apoptotic cysteine protease highly related to
 RT CPP32.";
 RL CANCER RES. 55:6045-6052(1995).
 RN [4]
 RP SEQUENCE FROM N.A. (ALPHA AND ALPHA' ISOFORMS).
 RC TISSUE-FETAL LUNG, AND FETAL SPLEEN;
 RX MEDLINE: 97224489.
 RA JUAN T.S.-C., MCNIECE I.K., ARGENTO J.M., JENKINS N.A., GILBERT D.J.,
 RA COPELAND N.G., FLETCHER F.A.;
 RT "Identification and mapping of Casp7, a cysteine protease resembling
 RT CPP32 beta, interleukin-1 beta converting enzyme, and CED-3.";
 RL GENOMICS 40:86-93(1997).
 RN [5]
 RP PROCESSING.
 RX MEDLINE: 96353838.
 RA FERNANDES-ALNEMRI T., ARMSTRONG R.C., KREBS J., SRINIVASULA S.M.,
 RA WANG L., BOLLRICH F., FRITZ L.C., TRAPANI J.A., TOMASELLI K.J.,
 RA LITWACK G., ALNEMRI E.S.;
 RT "In vitro activation of CPP32 and Mch3 by Mch4, a novel human
 RT apoptotic cysteine protease containing two FADP-like domains.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:7464-7469(1996).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES AND ACTIVATES STEROL
 CC REGULATORY ELEMENT BINDING PROTEINS (SREBPS). PROTEOLYTICALLY
 CC CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A 216-ASP-1-GLY-217
 CC BOND. OVEREXPRESSION PROMOTES PROGRAMMED CELL DEATH.
 CC -1- SUBUNIT: HETERODIMER OF A 20 KD (P20) AND A 11 KD (P11) SUBUNIT.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, SKELETAL MUSCLE,
 CC LIVER, KIDNEY, SPLEEN AND HEART, AND MODERATELY IN TESTIS. NO
 CC EXPRESSION IN THE BRAIN.
 CC -1- PTM: CLEAVAGES BY GRANZYME B OR CASPASE-10 GENERATE THE TWO ACTIVE
 CC SUBUNITS. PROPEPTIDE DOMAINS CAN ALSO BE CLEAVED EFFICIENTLY BY
 CC CPP32 PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
 CC CASPASE-7 AND THE LARGE SUBUNIT OF CPP32 PROTEASE ALSO OCCUR AND
 CC VICE VERSA.
 CC -1- ALTERNATIVE PRODUCTS: THREE ISOFORMS, ALPHA, BETA, AND ALPHA' ARE
 CC PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN HERE IS THAT
 CC OF ALPHA ISOFORM. THE BETA ISOFORM IS NOT PROTEOLYTICALLY ACTIVE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
 CC CASPASE FAMILY.
 CC -1- CAUTION: WHAT WE CALL 'ALPHA' ISOFORM IS KNOWN IN REF.4 AS BETA,
 CC BUT AS BETA IS ALREADY DEFINED IN REF.3 WE HAVE CALLED IT ALPHA'.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U39613; G1125073; -
 CC DR EMBL: U40281; G1167555; -
 CC DR EMBL: U37448; G1117847; -
 CC DR EMBL: U37449; G1117849; -
 CC DR EMBL: U67319; G1894913; -
 CC DR EMBL: U67320; G1894915; -
 CC DR MIM: 601761; -
 CC DR PROSITE: PS01121; CASPASE_HIS: 1.
 CC DR PROSITE: PS01122; CASPASE_CYS: 1.

827	5	1.0	339	1	Y4FO_RHISN	PUTATIVE ABC TRANSPORT	3.29e+03	900	5	1.0	512	1	YES6_YEAST	HYPOTHETICAL 56.6 KD P	3.29e+03		
828	5	1.0	341	1	VP3_GFLV	PROTEIN.	3.29e+03	901	5	1.0	512	1	YOKR_MYCTU	HYPOTHETICAL 53.6 KD P	3.29e+03		
829	5	1.0	342	1	YGIL_YEAST	HYPOTHETICAL 38.5 KD P	3.29e+03	902	5	1.0	519	1	VP6_WTAVU	STRUCTURAL PROTEIN P6	3.29e+03		
830	5	1.0	346	1	YOC2_CABEL	HYPOTHETICAL 39.2 KD P	3.29e+03	903	5	1.0	519	1	YOG1_CABEL	HYPOTHETICAL 59.2 KD P	3.29e+03		
831	5	1.0	346	1	YNP1_CABEL	HYPOTHETICAL 42.2 KD P	3.29e+03	904	5	1.0	543	1	ZYX_CHICK	ZYXIN.	3.29e+03		
832	5	1.0	348	1	XKDT_BACSU	PHAGE-LIKE ELEMENT PBS	3.29e+03	905	5	1.0	543	1	ZN08_HUMAN	ZINC FINGER PROTEIN 8	3.29e+03		
833	5	1.0	348	1	Z134_HUMAN	ZINC FINGER PROTEIN 13	3.29e+03	906	5	1.0	548	1	YDID_ECOLI	HYPOTHETICAL 60.8 KD P	3.29e+03		
834	5	1.0	349	1	Y028_BORBU	HYPOTHETICAL PROTEIN B	3.29e+03	907	5	1.0	551	1	YIDE_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03		
835	5	1.0	351	1	YIIG_ECOLI	HYPOTHETICAL 39.3 KD P	3.29e+03	908	5	1.0	559	1	YABC_SCHPO	PUTATIVE INORGANIC PHO	3.29e+03		
836	5	1.0	353	1	WN11_XENIA	WNT-11 PROTEIN PRECURS	3.29e+03	909	5	1.0	577	1	YGSU_YEAST	HYPOTHETICAL 67.3 KD P	3.29e+03		
837	5	1.0	359	1	Y4UE_RHISN	PUTATIVE TRANSPOSASE Y	3.29e+03	910	5	1.0	578	1	YAXB_SCHPO	PUTATIVE ATP-DEPENDENT	3.29e+03		
838	5	1.0	360	1	Y303_HELPY	PROBABLE GTP-BINDING P	3.29e+03	911	5	1.0	584	1	YMB3_YEAST	HYPOTHETICAL 66.8 KD P	3.29e+03		
839	5	1.0	360	1	YTPX_TTV1	VIRAL PROTEIN TPX.	3.29e+03	912	5	1.0	585	1	Y187_MYCCE	HYPOTHETICAL ABC TRANS	3.29e+03		
840	5	1.0	360	1	Y0FA_CABEL	HYPOTHETICAL 39.7 KD P	3.29e+03	913	5	1.0	597	1	KRP4_KILDA	HYPOTHETICAL KILLER PL	3.29e+03		
841	5	1.0	361	1	Y101_NPVAC	HYPOTHETICAL 41.5 KD P	3.29e+03	914	5	1.0	599	1	YGM5_YEAST	HYPOTHETICAL 68.5 KD P	3.29e+03		
842	5	1.0	361	1	Y4DE_SCHPO	HYPOTHETICAL 40.0 KD P	3.29e+03	915	5	1.0	599	1	YU01_YEAST	HYPOTHETICAL 68.5 KD P	3.29e+03		
843	5	1.0	366	1	Y433_ARCFU	HYPOTHETICAL PROTEIN A	3.29e+03	916	5	1.0	607	1	YSCC_YEREN	YOP PROTEINS TRANSLOCA	3.29e+03		
844	5	1.0	368	1	YAC3_SCHPO	HYPOTHETICAL 37.9 KD P	3.29e+03	917	5	1.0	614	1	Y3A3_SCHPO	HYPOTHETICAL 67.3 KD T	3.29e+03		
845	5	1.0	369	1	YOK3_CABEL	HYPOTHETICAL 41.7 KD P	3.29e+03	918	5	1.0	620	1	YED3_THEMA	HYPOTHETICAL 72.9 KD P	3.29e+03		
846	5	1.0	372	1	YJGP_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03	919	5	1.0	624	1	YED7_METUA	HYPOTHETICAL PROTEIN M	3.29e+03		
847	5	1.0	373	1	YXX1_ACHAM	HYPOTHETICAL PROTEIN (3.29e+03	920	5	1.0	628	1	YKPS_YEAST	HYPOTHETICAL 72.2 KD P	3.29e+03		
848	5	1.0	373	1	Y38K_THETE	HYPOTHETICAL 38 KD PRO	3.29e+03	921	5	1.0	631	1	YF14_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03		
849	5	1.0	374	1	YNOC_YEAST	HYPOTHETICAL 43.9 KD P	3.29e+03	922	5	1.0	632	1	YDIF_BACSU	HYPOTHETICAL ABC TRANS	3.29e+03		
850	5	1.0	375	1	Y4ED_RHISN	HYPOTHETICAL 42.2 KD P	3.29e+03	923	5	1.0	642	1	VP74_NPVCF	HYPOTHETICAL	3.29e+03		
851	5	1.0	377	1	Y1XN_BACSU	HYPOTHETICAL 42.9 KD P	3.29e+03	924	5	1.0	645	1	VPJ1_CABEL	HYPOTHETICAL 75.1 KD P	3.29e+03		
852	5	1.0	378	1	YK84_CABEL	HYPOTHETICAL 43.0 KD P	3.29e+03	925	5	1.0	654	1	YJHG_ECOLI	HYPOTHETICAL 70.0 KD P	3.29e+03		
853	5	1.0	379	1	YFG2_YEAST	HYPOTHETICAL 45.3 KD P	3.29e+03	926	5	1.0	655	1	Y390_MYCCE	HYPOTHETICAL ATP-BINDI	3.29e+03		
854	5	1.0	381	1	YN06_CABEL	HYPOTHETICAL 42.9 KD P	3.29e+03	927	5	1.0	660	1	Y390_MYCEN	HYPOTHETICAL ATP-BINDI	3.29e+03		
855	5	1.0	385	1	YCT2_BACFI	HYPOTHETICAL 43.4 KD P	3.29e+03	928	5	1.0	661	1	Y05L_MYCTU	HYPOTHETICAL 70.4 KD P	3.29e+03		
856	5	1.0	390	1	YFGB_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03	929	5	1.0	661	1	WH13_YEAST	WH13 PROTEIN.	3.29e+03		
857	5	1.0	391	1	Y4CE_RHISN	HYPOTHETICAL 43.4 KD P	3.29e+03	930	5	1.0	662	1	YF18_YEAST	HYPOTHETICAL 75.9 KD P	3.29e+03		
858	5	1.0	395	1	YCHE_CABEL	PUTATIVE GTP-BINDING P	3.29e+03	931	5	1.0	663	1	Y104_METUA	HYPOTHETICAL ATP-BINDI	3.29e+03		
859	5	1.0	400	1	YNJ2_YEAST	HYPOTHETICAL 45.5 KD P	3.29e+03	932	5	1.0	670	1	YER8_YEAST	HYPOTHETICAL 72.0 KD P	3.29e+03		
860	5	1.0	401	1	YD76_METUM	HYPOTHETICAL 43.7 KD P	3.29e+03	933	5	1.0	676	1	Y043_YEAST	HYPOTHETICAL 77.0 KD P	3.29e+03		
861	5	1.0	404	1	YJUN_YEAST	HYPOTHETICAL 46.4 KD P	3.29e+03	934	5	1.0	683	1	YK86_CABEL	HYPOTHETICAL 76.4 KD P	3.29e+03		
862	5	1.0	406	1	YD1B_SCHPO	PUTATIVE 2-ISOPROPYLMA	3.29e+03	935	5	1.0	683	1	YK87_YEAST	HYPOTHETICAL 78.3 KD P	3.29e+03		
863	5	1.0	407	1	Y503_METUA	HYPOTHETICAL 45.4 KD G	3.29e+03	936	5	1.0	689	1	YBP3_YEAST	HYPOTHETICAL 77.3 KD P	3.29e+03		
864	5	1.0	407	1	Y4CK_RHISN	PUTATIVE REPLICATION P	3.29e+03	937	5	1.0	702	1	YAI4_YEAST	HYPOTHETICAL 78.2 KD P	3.29e+03		
865	5	1.0	409	1	Y4DM_RHISN	HYPOTHETICAL 45.3 KD P	3.29e+03	938	5	1.0	715	1	Y070_MYCTU	HYPOTHETICAL 78.2 KD P	3.29e+03		
866	5	1.0	410	1	YD66_METTH	HYPOTHETICAL PROTEIN M	3.29e+03	939	5	1.0	722	1	Y290_HAEN	PROBABLE CATION-TRANSP	3.29e+03		
867	5	1.0	411	1	Y588_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03	940	5	1.0	724	1	YCF4_YEAST	HYPOTHETICAL 83.2 KD P	3.29e+03		
868	5	1.0	411	1	Y635_METUA	HYPOTHETICAL PROTEIN M	3.29e+03	941	5	1.0	725	1	YMT6_YEAST	HYPOTHETICAL 82.0 KD P	3.29e+03		
869	5	1.0	414	1	YCFW_ECOLI	HYPOTHETICAL 45.3 KD P	3.29e+03	942	5	1.0	728	1	Y376_METUA	HYPOTHETICAL PROTEIN M	3.29e+03		
870	5	1.0	414	1	YCHO_ECOLI	HYPOTHETICAL 47.0 KD P	3.29e+03	943	5	1.0	730	1	Y4E5_SCHPO	HYPOTHETICAL 79.3 KD P	3.29e+03		
871	5	1.0	417	1	Y146_MYCCE	HYPOTHETICAL PROTEIN M	3.29e+03	944	5	1.0	744	1	YGV4_HALSO	HYPOTHETICAL 80.2 KD P	3.29e+03		
872	5	1.0	424	1	Y4BN_RHISN	HYPOTHETICAL 47.6 KD P	3.29e+03	945	5	1.0	751	1	Y1X5_CABEL	HYPOTHETICAL 84.8 KD P	3.29e+03		
873	5	1.0	431	1	Y4BN_RHISN	HYPOTHETICAL 48.4 KD P	3.29e+03	946	5	1.0	751	1	YEHM_ECOLI	HYPOTHETICAL 83.4 KD P	3.29e+03		
874	5	1.0	431	1	Y4NC_SCHPO	HYPOTHETICAL ALDEHYDE-	3.29e+03	947	5	1.0	759	1	Y705_MYCTU	PROBABLE CATION-TRANSP	3.29e+03		
875	5	1.0	435	1	Y027_HUMAN	HYPOTHETICAL PROTEIN K	3.29e+03	948	5	1.0	772	1	Y1C1_ECOLI	DNA-REPAIR PROTEIN COM	3.29e+03		
876	5	1.0	438	1	YJBI_ECOLI	HYPOTHETICAL 51.7 KD P	3.29e+03	949	5	1.0	782	1	XPB_HUMAN	DNA-REPAIR PROTEIN B	3.29e+03		
877	5	1.0	442	1	Y4MB_RHISN	HYPOTHETICAL ZINC PROT	3.29e+03	950	5	1.0	783	1	XPB_MOUSE	TAIL TUBULAR PROTEIN B	3.29e+03		
878	5	1.0	447	1	Y329_MYCCE	HYPOTHETICAL 50.2 KD G	3.29e+03	951	5	1.0	794	1	YTPB_BPT7	DNA EXCISION REPAIR PR	3.29e+03		
879	5	1.0	448	1	VP1_CAVCI	PROBABLE COAT PROTEIN	3.29e+03	952	5	1.0	795	1	Y348_METUA	HYPOTHETICAL PROTEIN M	3.29e+03		
880	5	1.0	449	1	YEGQ_ECOLI	PUTATIVE PROTEASE IN B	3.29e+03	953	5	1.0	802	1	YGN9_YEAST	HYPOTHETICAL 90.8 KD P	3.29e+03		
881	5	1.0	453	1	ZXDA_HUMAN	ZINC FINGER X-LINKED P	3.29e+03	954	5	1.0	807	1	YCD5_ECOLI	HYPOTHETICAL 92.2 KD P	3.29e+03		
882	5	1.0	457	1	YV54_HSV6U	PROTEIN U54.	3.29e+03	955	5	1.0	807	1	YH42_YEAST	HYPOTHETICAL 96.8 KD P	3.29e+03		
883	5	1.0	458	1	Y086_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03	956	5	1.0	823	1	Y343_METUA	HYPOTHETICAL PROTEIN M	3.29e+03		
884	5	1.0	459	1	Y030_NPVOP	PUTATIVE PROTEASE H104	3.29e+03	957	5	1.0	849	1	Y07C_MYCTU	HYPOTHETICAL ABC TRANS	3.29e+03		
885	5	1.0	460	1	YEGQ_HAEN	HYPOTHETICAL PROTEIN M	3.29e+03	958	5	1.0	859	1	YG72_HAEN	HYPOTHETICAL PROTEIN H	3.29e+03		
886	5	1.0	460	1	Y061_MYCCE	HYPOTHETICAL 53.5 KD P	3.29e+03	959	5	1.0	881	1	YLB3_CABEL	HYPOTHETICAL 100.9 KD P	3.29e+03		
887	5	1.0	461	1	YEBD_SCHPO	HYPOTHETICAL 53.5 KD P	3.29e+03	960	5	1.0	892	1	YN86_YEAST	HYPOTHETICAL 104.5 KD P	3.29e+03		
888	5	1.0	462	1	Y01J_CABEL	PROBABLE AMINOTRANSFER	3.29e+03	961	5	1.0	924	1	YB53_YEAST	HYPOTHETICAL 103.2 KD	3.29e+03		
889	5	1.0	467	1	Y01J_CABEL	PROBABLE AMINOTRANSFER	3.29e+03	962	5	1.0	932	1	YV16_SCHPO	HYPOTHETICAL 105.3 KD	3.29e+03		
890	5	1.0	470	1	XNIF_XENIA	LOW MOLECULAR WEIGHT N	3.29e+03	963	5	1.0	950	1	YK15_HUMAN	HYPOTHETICAL PROTEIN K	3.29e+03		
891	5	1.0	472	1	YAHG_ECOLI	HYPOTHETICAL 50.4 KD P	3.29e+03	964	5	1.0	952	1	YNM7_YEAST	HYPOTHETICAL 105.8 KD	3.29e+03		
892	5	1.0	477	1	Y6U1_ECOLI	HYPOTHETICAL 52.1 KD P	3.29e+03	965	5	1.0	963	1	YD43_SCHPO	HYPOTHETICAL 111.9 KD	3.29e+03		
893	5	1.0	489	1	YEN5_YEAST	HYPOTHETICAL 54.6 KD P	3.29e+03	966	5	1.0	1002	1	YEMA_DROME	YEMANUCLEIN-ALPHA.	3.29e+03		
894	5	1.0	490	1	YPLC_CLOPE	HYPOTHETICAL 55.1 KD P	3.29e+03	967	5	1.0	1005	1	YCF1_OENBE	HYPOTHETICAL PROTEIN (3.29e+03		
895	5	1.0	491	1	YVGC_ECOLI	HYPOTHETICAL 55.3 KD P	3.29e+03	968	5	1.0							
896	5	1.0	497	1	YVGC_ECOLI	REGULATORY PROTEIN WET	3.29e+03	969	5	1.0							
897	5	1.0	499	1	WETA_PENCH	HYPOTHETICAL 57.9 KD P	3.29e+03	970	5	1.0							
898	5	1.0															

681	6	1.2	1316	1	RPOC_MYCTU	DNA-DIRECTED RNA POLYM	1.29e+02	754	5	1.0	148	1	YGEF_ECOLI	HYPOTHETICAL 16.9 KD P	3.29e+03
682	6	1.2	1317	1	RPOD_SYNY3	DNA-DIRECTED RNA POLYM	1.29e+02	755	5	1.0	148	1	YCT9_YEAST	HYPOTHETICAL 16.0 KD P	3.29e+03
683	6	1.2	1319	1	MNI_HUMAN	PROBABLE TUMOR SUPPRESS	1.29e+02	756	5	1.0	151	1	YD88_SCHPO	HYPOTHETICAL 17.5 KD P	3.29e+03
684	6	1.2	1332	1	XKDO_BACSU	PHAGE-LIKE ELEMENT PBS	1.29e+02	757	5	1.0	151	1	YD88_SYNY3	HYPOTHETICAL 17.3 KD P	3.29e+03
685	6	1.2	1358	1	PUR4_YEAST	PHOSPHORIBOSYLFORMYLIC	1.29e+02	758	5	1.0	155	1	YFH3_YEAST	HYPOTHETICAL 18.2 KD P	3.29e+03
686	6	1.2	1386	1	Y064_MYCPN	HYPOTHETICAL PROTEIN M	1.29e+02	759	5	1.0	166	1	YELI_SCHPO	HYPOTHETICAL PROTEIN C	3.29e+03
687	6	1.2	1396	1	VUTE_BPT5	L-SHAPED TAIL FIBRE PR	1.29e+02	760	5	1.0	158	1	Y097_MYCTU	HYPOTHETICAL 18.2 KD P	3.29e+03
688	6	1.2	1402	1	SALM_DROVI	HOMOTIC PROTEIN SPALT	1.29e+02	761	5	1.0	175	1	Y433_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
689	6	1.2	1403	1	NKCR_HUMAN	NK-TUMOR RECOGNITION P	1.29e+02	762	5	1.0	178	1	Y433_RHISN	HYPOTHETICAL 20.4 KD P	3.29e+03
690	6	1.2	1416	1	RPOC_TREPA	DNA-DIRECTED RNA POLYM	1.29e+02	763	5	1.0	189	1	YML6_PARTA	HYPOTHETICAL 22.4 KD P	3.29e+03
691	6	1.2	1518	1	KRKI_YEAST	PROBABLE SERINE/THREON	1.29e+02	764	5	1.0	191	1	Y645_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
692	6	1.2	1559	1	STCI_EMENT	POTATIVE SERTIGMOTICS	1.29e+02	765	5	1.0	192	1	YFMA_YEAST	HYPOTHETICAL 20.5 KD P	3.29e+03
693	6	1.2	1574	1	RPOC_AQUAE	DNA-DIRECTED RNA POLYM	1.29e+02	766	5	1.0	194	1	YFMA_SCHPO	HYPOTHETICAL 22.4 KD P	3.29e+03
694	6	1.2	1592	1	YHDS_YEAST	PROBABLE ATP-DEPENDENT	1.29e+02	767	5	1.0	195	1	YFEC_LACLC	HYPOTHETICAL PROTEIN I	3.29e+03
695	6	1.2	1603	1	PSC_DROME	POSTERIOR SEX COMBS PR	1.29e+02	768	5	1.0	199	1	YH33_MYCTU	HYPOTHETICAL 20.8 KD P	3.29e+03
696	6	1.2	1659	1	VIT_ONCMY	VITELLOGENIN PRECURSOR	1.29e+02	769	5	1.0	203	1	YBMA_YEAST	HYPOTHETICAL 22.6 KD P	3.29e+03
697	6	1.2	1676	1	APSA_EMENT	ANUCLEARE PRIMARY STER	1.29e+02	770	5	1.0	203	1	Y296_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
698	6	1.2	1683	1	YD9_YEAST	HYPOTHETICAL 191.5 KD	1.29e+02	771	5	1.0	209	1	YCR7_YEAST	HYPOTHETICAL RAS-RELA	3.29e+03
699	6	1.2	1698	1	Y076_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02	772	5	1.0	209	1	Y139_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
700	6	1.2	1722	1	RBB2_HUMAN	RETINOBLASTOMA BINDING	1.29e+02	773	5	1.0	210	1	Y1H8_ECOLI	HYPOTHETICAL GRP-BINDI	3.29e+03
701	6	1.2	1729	1	RBP5_YEAST	RNA BIOGENESIS PROTEI	1.29e+02	774	5	1.0	212	1	YME3_YEAST	HYPOTHETICAL 23.9 KD P	3.29e+03
702	6	1.2	1736	1	ZOI_HUMAN	TIGHT JUNCTION PROTEIN F	1.29e+02	775	5	1.0	213	1	YH06_YEAST	HYPOTHETICAL 24.6 KD P	3.29e+03
703	6	1.2	1745	1	YLS1_CAEEL	HYPOTHETICAL PROTEIN F	1.29e+02	776	5	1.0	218	1	YTXE_BACME	HYPOTHETICAL 24.6 KD P	3.29e+03
704	6	1.2	1765	1	LMH1_HUMAN	LAMININ BETA-1 CHAIN N	1.29e+02	777	5	1.0	219	1	Y556_SYNY3	HYPOTHETICAL 24.0 KD P	3.29e+03
705	6	1.2	1805	1	HMW2_MYCCE	CYTADHERENCE HIGH MOLE	1.29e+02	778	5	1.0	225	1	YGLX_ECOLI	PROBABLE TRANSCRIPTION	3.29e+03
706	6	1.2	1818	1	HMW2_MYCPN	CYTADHERENCE HIGH MOLE	1.29e+02	779	5	1.0	225	1	YXN1_EMENT	ENDO-1,4-BETA-XYLANASE	3.29e+03
707	6	1.2	2194	1	SC16_YEAST	MULTIDOMAIN VESICLE CO	1.29e+02	780	5	1.0	226	1	YCKA_BACSU	PROBABLE AMINO-ACID AB	3.29e+03
708	6	1.2	2209	1	Y166_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02	781	5	1.0	236	1	Y247_MYCCE	HYPOTHETICAL PROTEIN M	3.29e+03
709	6	1.2	2252	1	POL1_GCMV	RNAL POLYPROTEIN (250	1.29e+02	782	5	1.0	242	1	YBGP_ECOLI	HYPOTHETICAL FIBRILP	3.29e+03
710	6	1.2	2264	1	POL1_TBRVS	CHONDROTTIN SULFATE PR	1.29e+02	783	5	1.0	245	1	YBPA_BURCE	HYPOTHETICAL TRANSCRIP	3.29e+03
711	6	1.2	2325	1	FRAP_HUMAN	FKBP-RAPAMYCIN ASSOCIA	1.29e+02	784	5	1.0	247	1	Y076_MYCTU	POTATIVE OXIDOREDUCTAS	3.29e+03
712	6	1.2	2349	1	FRAP_HUMAN	FKBP-RAPAMYCIN ASSOCIA	1.29e+02	785	5	1.0	249	1	YH85_YEAST	HYPOTHETICAL 27.6 KD P	3.29e+03
713	6	1.2	2349	1	FRAP_HUMAN	FKBP-RAPAMYCIN ASSOCIA	1.29e+02	786	5	1.0	249	1	YF57_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
714	6	1.2	2394	1	TEGU_DROVI	SEVENTEEN PROTEIN (EC	1.29e+02	787	5	1.0	251	1	Y4EF_RHISN	POTATIVE INTEGRASE/REC	3.29e+03
715	6	1.2	2763	1	Y166_HUMAN	LARGE TEGUMENT PROTEIN	1.29e+02	788	5	1.0	255	1	YHJH_ECOLI	HYPOTHETICAL 29.6 KD P	3.29e+03
716	6	1.2	3075	1	LMH1_HUMAN	LAMININ ALPHA-1 CHAIN	1.29e+02	789	5	1.0	261	1	Y076_MYCTU	HYPOTHETICAL TRANSCRIP	3.29e+03
717	6	1.2	3084	1	LMH1_MOUSE	LAMININ ALPHA-1 CHAIN	1.29e+02	790	5	1.0	261	1	Y046_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
718	6	1.2	3144	1	HD_HUMAN	HUNTINGTIN (HUNTINGTON	1.29e+02	791	5	1.0	262	1	YGB6_YEAST	HYPOTHETICAL 29.4 KD P	3.29e+03
719	6	1.2	3175	1	RPOA_EAV	POL POLYPROTEIN (ORF1A	1.29e+02	792	5	1.0	264	1	YHFI_SALTY	HYPOTHETICAL 26.3 KD P	3.29e+03
720	6	1.2	3386	1	POLG_DEN4	GENOME POLYPROTEIN (CO	1.29e+02	793	5	1.0	264	1	YHFI_SALTY	HYPOTHETICAL 26.3 KD P	3.29e+03
721	6	1.2	3519	1	OL56_STRAT	OLEANDROMYCIN POLKETID	1.29e+02	794	5	1.0	269	1	YBBO_ECOLI	HYPOTHETICAL OXIDOREDU	3.29e+03
722	6	1.2	3924	1	ANKK_HUMAN	ANKRYNIN BRAIN VARIANT	1.29e+02	795	5	1.0	269	1	YDLO_SCHPO	HYPOTHETICAL 29.9 KD P	3.29e+03
723	6	1.2	4393	1	PGSM_HUMAN	BASEMENT MEMBRANE-SPEC	1.29e+02	796	5	1.0	270	1	YDLO_SCHPO	HYPOTHETICAL 31.5 KD P	3.29e+03
724	6	1.2	4427	1	PXSL_BACSU	POTATIVE POLYKETIDE SY	1.29e+02	797	5	1.0	275	1	YD84_YEAST	HYPOTHETICAL 30.0 KD P	3.29e+03
725	6	1.2	4563	1	APB_HUMAN	APOLIPOPROTEIN B-100 P	1.29e+02	798	5	1.0	279	1	Y284_CHVP1	HYPOTHETICAL PROTEIN A	3.29e+03
726	5	1.0	57	1	YCI8_AGLNE	HYPOTHETICAL 7.0 KD PR	3.29e+03	799	5	1.0	282	1	Y32K_BNYVG	RNA-4 HYPOTHETICAL 31.	3.29e+03
727	5	1.0	65	1	YCF9_CYAPA	HYPOTHETICAL 6.9 KD PR	3.29e+03	800	5	1.0	286	1	YMD3_CAEEL	HYPOTHETICAL 32.1 KD P	3.29e+03
728	5	1.0	77	1	YEDF_ECOLI	HYPOTHETICAL 8.6 KD PR	3.29e+03	801	5	1.0	286	1	Y137_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
729	5	1.0	84	1	YB96_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	802	5	1.0	287	1	YCX1_METJA	HYPOTHETICAL 34.5 KD P	3.29e+03
730	5	1.0	88	1	YGTU_ECOLI	HYPOTHETICAL 11.2 KD P	3.29e+03	803	5	1.0	291	1	Y32K_SSV1	HYPOTHETICAL 31.7 KD P	3.29e+03
731	5	1.0	98	1	Y4CB_RHISN	HYPOTHETICAL 10.2 KD P	3.29e+03	804	5	1.0	291	1	YG47_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03
732	5	1.0	99	1	Y073_NPVAC	HYPOTHETICAL 11.5 KD P	3.29e+03	805	5	1.0	293	1	YUCO_ECOLI	HYPOTHETICAL ABC TRANS	3.29e+03
733	5	1.0	102	1	YK85_CAEEL	HYPOTHETICAL 11.7 KD P	3.29e+03	806	5	1.0	293	1	YNEJ_ECOLI	HYPOTHETICAL TRANSCRIP	3.29e+03
734	5	1.0	108	1	YJ79_YEAST	HYPOTHETICAL 12.5 KD P	3.29e+03	807	5	1.0	294	1	YU51_HSVJ7	G-PROTEIN COUPLED RECE	3.29e+03
735	5	1.0	110	1	Y18K_MSVK	HYPOTHETICAL 12.7 KD P	3.29e+03	808	5	1.0	294	1	YFEC_YERPE	CHELATED IRON TRANSPOR	3.29e+03
736	5	1.0	111	1	YKXD_PORPU	HYPOTHETICAL 13.2 KD P	3.29e+03	809	5	1.0	298	1	YDHE_ECOLI	HYPOTHETICAL OXIDOREDU	3.29e+03
737	5	1.0	112	1	Y122_CAEEL	HYPOTHETICAL 12.2 KD P	3.29e+03	810	5	1.0	299	1	YED5_YEAST	HYPOTHETICAL 34.8 KD P	3.29e+03
738	5	1.0	114	1	Y126_ADE07	HYPOTHETICAL 12.6 KD P	3.29e+03	811	5	1.0	303	1	VP40_MABVP	MATRIX PROTEIN VP40 (V	3.29e+03
739	5	1.0	121	1	YG47_YEAST	HYPOTHETICAL 13.4 KD P	3.29e+03	812	5	1.0	306	1	YDS4_SCHPO	HYPOTHETICAL 35.8 KD P	3.29e+03
740	5	1.0	122	1	Y102_NPVAC	HYPOTHETICAL 13.3 KD P	3.29e+03	813	5	1.0	310	1	YVAH_ECOLI	HYPOTHETICAL TRANSCRIP	3.29e+03
741	5	1.0	122	1	YD1C_ECOLI	HYPOTHETICAL 13.3 KD P	3.29e+03	814	5	1.0	313	1	YFEM_MTCOL	HYPOTHETICAL PROTEIN I	3.29e+03
742	5	1.0	123	1	WNT7_EYATR	WNT-7 PROTEIN (FRAGMEN	3.29e+03	815	5	1.0	315	1	YLX1_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
743	5	1.0	123	1	WNT3A_ALOUV	WNT-3A PROTEIN (FRAGME	3.29e+03	816	5	1.0	315	1	YK93_CAEEL	HYPOTHETICAL CHOLINE K	3.29e+03
744	5	1.0	125	1	Y02D_BPT4	HYPOTHETICAL 14.1 KD P	3.29e+03	817	5	1.0	316	1	YEEY_ECOLI	HYPOTHETICAL TRANSCRIP	3.29e+03
745	5	1.0	137	1	Y0G4_YEAST	HYPOTHETICAL 13.9 KD P	3.29e+03	818	5	1.0	317	1	YCHI_THEMEA	HYPOTHETICAL 36.4 KD P	3.29e+03
746	5	1.0	137	1	Y0HC_MYCTU	HYPOTHETICAL 15.0 KD P	3.29e+03	819	5	1.0	318	1	YDKE_ECOLI	HYPOTHETICAL 37.5 KD P	3.29e+03
747	5	1.0	140	1	YJ14_YEAST	HYPOTHETICAL 15.7 KD P	3.29e+03	820	5	1.0	324	1	YB61_YEAST	HYPOTHETICAL 37.5 KD P	3.29e+03
748	5	1.0	142	1	YFAT_SCHPO	HYPOTHETICAL PROTEIN H	3.29e+03	821	5	1.0	325	1	VP6_BTV11	VP6 PROTEIN (MINOR INN	3.29e+03
749	5	1.0	144	1	YB50_HAEIN	HYPOTHETICAL PROTEIN C	3.29e+03	822	5	1.0	325	1	VP6_BTV13	VP6 PROTEIN (MINOR INN	3.29e+03
750	5	1.0	146	1	YVJ1_YEAST	HYPOTHETICAL 17.2 KD P	3.29e+03	823	5	1.0	330	1	Y4KL_RHISN	HYPOTHETICAL 37.6 KD P	3.29e+03
751	5	1.0	147	1	YVJ1_YEAST	VERY HYPOTHETICAL 16.8	3.29e+03	824	5	1.0	332	1	Y0Q1_CAEEL	HYPOTHETICAL 37.5 KD P	3.29e+03
752	5	1.0	148	1	YERO_YEAST	HYPOTHETICAL 15.0 KD P	3.29e+03	825	5	1.0	336	1	YVJ3_CAEEL	HYPOTHETICAL 38.9 KD P	3.29e+03
753	5	1.0	148	1	YFBI2_YEAST	HYPOTHETICAL 17.2 KD P	3.29e+03	826	5	1.0	338	1	YABT_BACSU	PROBABLE SERINE/THREON	3.29e+03

535	6	1.2	595	1	YPD7_CAEEL	HYPOTHETICAL 66.7 KD P	1.29e+02	608	6	1.2	840	1	CC16_YEAST	CELL DIVISION CONTROL	1.29e+02
536	6	1.2	599	1	TYPA_HELPY	GTP-BINDING PROTEIN TY	1.29e+02	609	6	1.2	843	1	CYPL_BROMA	PEPTIDYLPROLYL ISOMERA	1.29e+02
537	6	1.2	601	1	OAR_DROME	OCTOPAMINE RECEPTOR PR	1.29e+02	610	6	1.2	853	1	PBPA_HAEIN	PENICILLIN-BINDING PRO	1.29e+02
538	6	1.2	603	1	PLK1_HUMAN	SERINE/THREONINE-PROTE	1.29e+02	611	6	1.2	859	1	ENV_ETIAY2	ENV POLYPROTEIN PRECUR	1.29e+02
539	6	1.2	603	1	PRIM_BACSU	DNA PRIMASE (EC 2.7.7.	1.29e+02	612	6	1.2	859	1	ENV_ETIAY1	ENV POLYPROTEIN PRECUR	1.29e+02
540	6	1.2	603	1	PLK1_MOUSE	SERINE/THREONINE-PROTE	1.29e+02	613	6	1.2	859	1	ENV_ETIAYC	ENV POLYPROTEIN PRECUR	1.29e+02
541	6	1.2	603	1	LEPA_SYNY3	GTP-BINDING PROTEIN LE	1.29e+02	614	6	1.2	859	1	ENV_ETIAY9	ENV POLYPROTEIN PRECUR	1.29e+02
542	6	1.2	605	1	AOR_PRRPU	TUNGSTEN-CONTAINING AL	1.29e+02	615	6	1.2	859	1	ENV_ETIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
543	6	1.2	607	1	DB10_NICSY	RNA HELICASE-LIKE PROT	1.29e+02	616	6	1.2	859	1	ENV_ETIAYW	ENV POLYPROTEIN PRECUR	1.29e+02
544	6	1.2	609	1	YA58_METUA	HYPOTHETICAL PROTEIN M	1.29e+02	617	6	1.2	859	1	ENV_ETIAY3	ENV POLYPROTEIN PRECUR	1.29e+02
545	6	1.2	614	1	RI17_HUMAN	U1 SMALL NUCLEAR RIBON	1.29e+02	618	6	1.2	860	1	ENV_ETIAY5	ENV POLYPROTEIN PRECUR	1.29e+02
546	6	1.2	617	1	HEMA_MEASI	HEMAGGLUTININ-NEURAMIN	1.29e+02	619	6	1.2	868	1	CLP_TGRB1	HEAT SHOCK PROTEIN 100	1.29e+02
547	6	1.2	617	1	PEPS1_PSEAE	PYOCIN S1 (EC 3.1.21.1	1.29e+02	620	6	1.2	868	1	MIS5_SCHPO	MISS PROTEIN	1.29e+02
548	6	1.2	619	1	OM70_NEUCR	MITOCHONDRIAL PRECURSO	1.29e+02	621	6	1.2	868	1	MIB0_YEAST	NIP80 PROTEIN (NIP100	1.29e+02
549	6	1.2	622	1	YRT1_CAEEL	HYPOTHETICAL 71.6 KD P	1.29e+02	622	6	1.2	869	1	RIR3_YEAST	RIBONUCLEOSIDE-DIPHOSP	1.29e+02
550	6	1.2	622	1	YAG1_SCHPO	HYPOTHETICAL PROTEIN C	1.29e+02	623	6	1.2	875	1	LON1_SPTOL	MITOCHONDRIAL ION PROT	1.29e+02
551	6	1.2	630	1	TNPB_STAAU	TRANSPPOSASE B (TRANSPO	1.29e+02	624	6	1.2	879	1	MYSF_DROME	PARMAYOSIN, LONG FORM.	1.29e+02
552	6	1.2	630	1	TNPE_STAAU	TRANSPPOSASE B (TRANSPO	1.29e+02	625	6	1.2	885	1	LON1_MAIZE	MITOCHONDRIAL ION PROT	1.29e+02
553	6	1.2	632	1	YK55_CAEEL	HYPOTHETICAL 71.9 KD P	1.29e+02	626	6	1.2	891	1	POL2_BAMMN	GENOME POLYPROTEIN 2	1.29e+02
554	6	1.2	637	1	YPOD_SCHPO	HYPOTHETICAL 71.6 KD P	1.29e+02	627	6	1.2	892	1	TRAI_STRFR	TRANSPPOSASE FOR TRANSP	1.29e+02
555	6	1.2	638	1	PAN1_RAT	TRANSCRIPTIONAL REGULA	1.29e+02	628	6	1.2	905	1	GYRA_RICPR	DNA GYRASE SUBUNIT A	1.29e+02
556	6	1.2	639	1	PRIM_MYCTU	DNA PRIMASE (EC 2.7.7.	1.29e+02	629	6	1.2	907	1	NDOG_ECOLI	NADH DEHYDROGENASE I C	1.29e+02
557	6	1.2	646	1	UYRB_METTH	EXCINUCLEASE ABC SUBUN	1.29e+02	630	6	1.2	907	1	NDOG_SALTU	NADH DEHYDROGENASE I C	1.29e+02
558	6	1.2	649	1	BCAL_MOUSE	BETA-GALACTOSIDASE PRE	1.29e+02	631	6	1.2	917	1	UGLB_HSVB2	GLYCOPROTEIN B-1 PRECU	1.29e+02
559	6	1.2	649	1	PAN2_RAT	TRANSCRIPTIONAL REGULA	1.29e+02	632	6	1.2	921	1	RB_MOUSE	RETINOBLASTOMA-ASSOCIA	1.29e+02
560	6	1.2	652	1	PAS1_DROME	FASCICLIN I PRECURSOR	1.29e+02	633	6	1.2	928	1	ATCL_YARLI	CALCIUM-TRANSPORTING A	1.29e+02
561	6	1.2	663	1	GR78_SCHPO	78 KD GLUCOSE REGULATE	1.29e+02	634	6	1.2	928	1	RB_HUMAN	RETINOBLASTOMA-ASSOCIA	1.29e+02
562	6	1.2	668	1	YBFI_YEAST	HYPOTHETICAL 73.8 KD P	1.29e+02	635	6	1.2	935	1	YJ31_YEAST	HYPOTHETICAL 108.4 KD	1.29e+02
563	6	1.2	669	1	PEPS1_YEAST	GLYCEROL UPTAKE/EFFLUX	1.29e+02	636	6	1.2	962	1	GUNA_PSEFL	ENDOLGLUCANASE A PRECUR	1.29e+02
564	6	1.2	673	1	YNN5_YEAST	HYPOTHETICAL 73.8 KD P	1.29e+02	637	6	1.2	979	1	PTPN_MOUSE	PROTEIN-TYROSINE PHOSP	1.29e+02
565	6	1.2	673	1	KR74_HSV11	GENE 74 PROTEIN KINASE	1.29e+02	638	6	1.2	979	1	PTPN_HUMAN	PROTEIN-TYROSINE PHOSP	1.29e+02
566	6	1.2	676	1	UI06_HSV11	VIRION PROTEIN UI6.	1.29e+02	639	6	1.2	983	1	PTPN_RAT	PROTEIN-TYROSINE PHOSP	1.29e+02
567	6	1.2	677	1	FLHA_BACSU	FLAGELLAR BIOSYNTHESIS	1.29e+02	640	6	1.2	1020	1	NFH_HUMAN	NEUROFILAMENT TRIPLET	1.29e+02
568	6	1.2	681	1	TRF_MANSE	TRANSFERRIN PRECURSOR.	1.29e+02	641	6	1.2	1021	1	MANA_ROMR	MANANN ENDO-1,4-BETA-M	1.29e+02
569	6	1.2	682	1	PILI_PSEAE	PILI PROTEIN.	1.29e+02	642	6	1.2	1030	1	PEX6_YEAST	PEROXISOME BIOSYNTHESI	1.29e+02
570	6	1.2	685	1	SNMA_DICD1	SNMA PROTEIN.	1.29e+02	643	6	1.2	1042	1	SUMA_DROME	SUPPRESSOR OF WHITE AP	1.29e+02
571	6	1.2	690	1	ATKB_SYNY3	POTASSIUM-TRANSPORTING	1.29e+02	644	6	1.2	1046	1	YJ14_YEAST	HYPOTHETICAL 117.2 KD	1.29e+02
572	6	1.2	691	1	HSF_DROME	HEAT SHOCK FACTOR PROT	1.29e+02	645	6	1.2	1065	1	CERU_HUMAN	CERULOPLASMIN PRECURSO	1.29e+02
573	6	1.2	695	1	XET_HUMAN	PROTEIN XET.	1.29e+02	646	6	1.2	1073	1	Y126_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02
574	6	1.2	700	1	PALY_WHEAT	PHENYLALANINE AMMONIA-	1.29e+02	647	6	1.2	1078	1	C27A_MOUSE	MHC CLASS II TRANSACTIN	1.29e+02
575	6	1.2	702	1	CCEM_HUMAN	CARCINOMERYONIC ANTIG	1.29e+02	648	6	1.2	1078	1	GYRB_SYNY3	DNA HELICASE SUBUNIT B	1.29e+02
576	6	1.2	702	1	COAT_SMV1	COAT PROTEIN (CAPSID P	1.29e+02	649	6	1.2	1081	1	UL52_HSVB	DNA HELICASE/PRIMASE C	1.29e+02
577	6	1.2	705	1	YWJE_BACSU	HYPOTHETICAL 79.2 KD P	1.29e+02	650	6	1.2	1086	1	NNTM_BOVIN	NAD(P) TRANSHYDROGENAS	1.29e+02
578	6	1.2	707	1	DCOR_LEIDO	ORNITHINE DECARBOXYLAS	1.29e+02	651	6	1.2	1087	1	NNTM_MOUSE	NEUROFILAMENT TRIPLET	1.29e+02
579	6	1.2	708	1	Y253_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02	652	6	1.2	1091	1	DIA_DROME	DIAPHANOUS PROTEIN.	1.29e+02
580	6	1.2	722	1	MASZ_ECOLI	MALATE SYNTHASE G (EC	1.29e+02	653	6	1.2	1092	1	DPD_DROME	DNA POLYMERASE DELTA C	1.29e+02
581	6	1.2	727	1	IF2M_BOVIN	TRANSLATION INITIATION	1.29e+02	654	6	1.2	1093	1	AF17_HUMAN	AF-17 PROTEIN.	1.29e+02
582	6	1.2	731	1	TOP1_TREPA	DNA TOPOISOMERASE I (E	1.29e+02	655	6	1.2	1097	1	IMB3_HUMAN	IMPORTIN BETA-3 SUBUNIT	1.29e+02
583	6	1.2	738	1	MASY_CORGL	MALATE SYNTHASE (EC 4.	1.29e+02	656	6	1.2	1111	1	YJBP_HAEIN	HYPOTHETICAL PROTEIN H	1.29e+02
584	6	1.2	739	1	UI47_HSVB	80.7 KD ALPHA TRANS-IN	1.29e+02	657	6	1.2	1128	1	PHX3_AVEA	PHYTOCHROME A TYPE 3	1.29e+02
585	6	1.2	741	1	MASY_MYCTU	PROBABLE MALATE SYNTHA	1.29e+02	658	6	1.2	1128	1	PHX3_AVEA	PHYTOCHROME A TYPE 3	1.29e+02
586	6	1.2	742	1	ZFA_MOUSE	ZINC FINGER AUTOSOMAL	1.29e+02	659	6	1.2	1130	1	C27A_HUMAN	MHC CLASS II TRANSACTIN	1.29e+02
587	6	1.2	742	1	UI47_HSVB	80.7 KD ALPHA TRANS-IN	1.29e+02	660	6	1.2	1132	1	BAT3_HUMAN	LARGE PROLINE-RICH PRO	1.29e+02
588	6	1.2	743	1	YK47_YEAST	HYPOTHETICAL 83.6 KD P	1.29e+02	661	6	1.2	1132	1	NUT1_YEAST	NUT1 PROTEIN.	1.29e+02
589	6	1.2	747	1	KE3B_MOUSE	KINESIN-LIKE PROTEIN K	1.29e+02	662	6	1.2	1139	1	SRE2_CRIGR	STEROL REGULATORY ELEM	1.29e+02
590	6	1.2	747	1	KE3B_MOUSE	KINESIN-LIKE PROTEIN K	1.29e+02	663	6	1.2	1141	1	SRE2_HUMAN	STEROL REGULATORY ELEM	1.29e+02
591	6	1.2	747	1	CI04_MOUSE	CHLORIDE CHANNEL PROTE	1.29e+02	664	6	1.2	1142	1	PAK1_YEAST	SERINE/THREONINE-PROTE	1.29e+02
592	6	1.2	749	1	SWAP_CAEEL	SMAP PROTEIN (SUPPRESS	1.29e+02	665	6	1.2	1184	1	POL2_GFLV	RNA2 POLYPROTEIN (131	1.29e+02
593	6	1.2	752	1	EFGM_RAT	ELONGATION FACTOR G, M	1.29e+02	666	6	1.2	1184	1	POC2_BACSU	DNA-DIRECTED RNA POLYM	1.29e+02
594	6	1.2	758	1	CHUD_TOBAC	MAGNESIUM-CHELATASE SU	1.29e+02	667	6	1.2	1209	1	DNB1_HSVB	MAJOR DNA-BINDING PROT	1.29e+02
595	6	1.2	760	1	DBP4_MOUSE	DIPEPTIDYL PEPTIDASE I	1.29e+02	668	6	1.2	1222	1	YNN5_CAEEL	HYPOTHETICAL 136.3 KD	1.29e+02
596	6	1.2	770	1	LEU2_CANNA	3-ISOPROPYLMALATE DEHY	1.29e+02	669	6	1.2	1224	1	RPOD_PORPU	DNA-DIRECTED RNA POLYM	1.29e+02
597	6	1.2	773	1	PCO2_PSES3	PENICILLIN ACYLASE II	1.29e+02	670	6	1.2	1228	1	SLAP_BACST	S-LAYER PROTEIN PRECUR	1.29e+02
598	6	1.2	773	1	FE0B_ECOLI	FERROUS IRON TRANSPORT	1.29e+02	671	6	1.2	1254	1	DPQG_YEAST	DNA POLYMERASE GAMMA	1.29e+02
599	6	1.2	776	1	VP42_ROTSS	OUTER CAPSID PROTEIN V	1.29e+02	672	6	1.2	1265	1	RPOD_YEAST	DNA-DIRECTED RNA POLYM	1.29e+02
600	6	1.2	776	1	VP42_ROTSS	OUTER CAPSID PROTEIN V	1.29e+02	673	6	1.2	1272	1	UBP2_YEAST	UBIQUITIN CARBOXYL-TER	1.29e+02
601	6	1.2	781	1	YB68_SCHPO	DYNAMIN-LIKE PROTEIN K	1.29e+02	674	6	1.2	1286	1	SK12_YEAST	ANTIVIRAL PROTEIN SK12	1.29e+02
602	6	1.2	781	1	KRAF_DROME	RAF HOMOLOG SERINE/THR	1.29e+02	675	6	1.2	1286	1	RPOD_GUTTH	DNA-DIRECTED RNA POLYM	1.29e+02
603	6	1.2	798	1	YLFN_CAEEL	HYPOTHETICAL 89.8 KD P	1.29e+02	676	6	1.2	1289	1	VG34_BPT4	LARGE TAIL FIBER PROTE	1.29e+02
604	6	1.2	812	1	PLMN_BOVIN	PLASMINOGEN PRECURSOR	1.29e+02	677	6	1.2	1290	1	RPOC_MYCPN	DNA-DIRECTED RNA POLYM	1.29e+02
605	6	1.2	815	1	ERK5_HUMAN	EXTRACELLULAR SIGNAL-R	1.29e+02	678	6	1.2	1292	1	RPOC_MYCPN	DNA-DIRECTED RNA POLYM	1.29e+02
606	6	1.2	824	1	PML3_HUMAN	PROBABLE TRANSCRIPTION	1.29e+02	679	6	1.2	1299	1	VP75_HSVB	PROBABLE MEMBRANE ANTI	1.29e+02
607	6	1.2	832	1	KIPI_SCHPO	KINESIN-LIKE PROTEIN 1	1.29e+02	680	6	1.2	1316	1	RPOC_MYCLE	DNA-DIRECTED RNA POLYM	1.29e+02

389	6	1.2	416	1	CP11_RAT	CONTRAPASIN-LIKE PROTEA	1.29e+02	462	6	1.2	500	1	PTG1_BOVIN	PROSTACYCLIN SYNTHASE	1.29e+02
390	6	1.2	419	1	ST5_MYCTU	SERYL-TRNA SYNTHETASE	1.29e+02	463	6	1.2	501	1	DHE3_BOVIN	GLUTAMATE DEHYDROGENAS	1.29e+02
391	6	1.2	420	1	CRFR_CHICK	CORTICOTROPIN RELEASEIN	1.29e+02	464	6	1.2	502	1	SCRB_HELPE	SUCROSE-6-PHOSPHATE HY	1.29e+02
392	6	1.2	420	1	MLIC_XENLA	MELATONIN RECEPTOR TYP	1.29e+02	465	6	1.2	501	1	Y190_HELPE	HYPOTHETICAL PROTEIN H	1.29e+02
393	6	1.2	422	1	PAX6_XENLA	PAIRED BOX PROTEIN PAX	1.29e+02	466	6	1.2	503	1	YOM5_CAEEL	HYPOTHETICAL 57.1 KD P	1.29e+02
394	6	1.2	422	1	PAX6_HUMAN	PAIRED BOX PROTEIN PAX	1.29e+02	467	6	1.2	503	1	DHE3_CHICK	GLUTAMATE DEHYDROGENAS	1.29e+02
395	6	1.2	422	1	PAX6_MOUSE	PAIRED BOX PROTEIN PAX	1.29e+02	468	6	1.2	505	1	MARK_SAXIN	PROBABLE INTRON MATURA	1.29e+02
396	6	1.2	422	1	PAX6_MOUSE	PAIRED BOX PROTEIN PAX	1.29e+02	469	6	1.2	506	1	KPY2_YEAST	PROBABLE KINASE 2 (PC	1.29e+02
397	6	1.2	423	1	PAX6_MOUSE	PAIRED BOX PROTEIN PAX	1.29e+02	470	6	1.2	508	1	TY3H_DROME	TYROSINE 3-MONOXYGENA	1.29e+02
398	6	1.2	423	1	FE1A_ARCFU	ELONGATION FACTOR 1-AL	1.29e+02	471	6	1.2	509	1	ATP0_MAIZE	ATP SYNTHASE ALPHA CHA	1.29e+02
399	6	1.2	423	1	IRK6_HUMAN	G PROTEIN-ACTIVATED IN	1.29e+02	472	6	1.2	508	1	VP67_NPVCF	MAJOR ENVELOPE GLYCOPR	1.29e+02
400	6	1.2	424	1	ZP3A_CALSO	ZONA PELLICULA SPERM-B	1.29e+02	473	6	1.2	509	1	VE2_HEV36	REGULATORY PROTEIN E2	1.29e+02
401	6	1.2	425	1	IRK6_MOUSE	G PROTEIN-ACTIVATED IN	1.29e+02	474	6	1.2	510	1	YICM_ECOLI	HYPOTHETICAL ABC TRANS	1.29e+02
402	6	1.2	425	1	IRK6_MOUSE	G PROTEIN-ACTIVATED IN	1.29e+02	475	6	1.2	512	1	VP67_NPVAC	MAJOR ENVELOPE GLYCOPR	1.29e+02
403	6	1.2	425	1	IRK6_MOUSE	G PROTEIN-ACTIVATED IN	1.29e+02	476	6	1.2	513	1	ST5_METTH	SERYL-TRNA SYNTHETASE	1.29e+02
404	6	1.2	426	1	WCAR_ECOLI	G PROTEIN-ACTIVATED IN	1.29e+02	477	6	1.2	514	1	ALAB_MOUSE	ALPHA-1B ADRENERGIC RE	1.29e+02
405	6	1.2	427	1	GBP2_YEAST	SINGLE-STRAND TLOMERERI	1.29e+02	478	6	1.2	515	1	ENV_BLVAV	ENV POLYPROTEIN PRECUR	1.29e+02
406	6	1.2	428	1	SYS_STAAT	SERYL-TRNA SYNTHETASE	1.29e+02	479	6	1.2	515	1	ENV_BLVAV	ENV POLYPROTEIN PRECUR	1.29e+02
407	6	1.2	433	1	LE21_ARCFU	PUTATIVE 3-ISOPROPYLMA	1.29e+02	480	6	1.2	515	1	ALAB_MESAU	ALPHA-1B ADRENERGIC RE	1.29e+02
408	6	1.2	435	1	YCM9_YEAST	HYPOTHETICAL 48.5 KD P	1.29e+02	481	6	1.2	515	1	ENV_BLV	ENV POLYPROTEIN PRECUR	1.29e+02
409	6	1.2	441	1	VE2_HEV38	REGULATORY PROTEIN E2	1.29e+02	482	6	1.2	515	1	ENV_BLVJ	ENV POLYPROTEIN PRECUR	1.29e+02
410	6	1.2	442	1	HRDB_STRCO	RNA POLYMERASE PRINCIP	1.29e+02	483	6	1.2	515	1	ENV_BLVAF	ENV POLYPROTEIN PRECUR	1.29e+02
411	6	1.2	444	1	CRFR_HUMAN	CORTICOTROPIN RELEASEIN	1.29e+02	484	6	1.2	515	1	ENV_BLV2	ENV POLYPROTEIN PRECUR	1.29e+02
412	6	1.2	445	1	BPI_RABIT	BACTERICIDAL PERMEABIL	1.29e+02	485	6	1.2	515	1	ALAB_RAT	ALPHA-1B ADRENERGIC RE	1.29e+02
413	6	1.2	448	1	OCT6_HUMAN	OCTAMER-BINDING TRANS	1.29e+02	486	6	1.2	515	1	VAC1_YEAST	VAC1 PROTEIN	1.29e+02
414	6	1.2	449	1	OCT6_MOUSE	OCTAMER-BINDING TRANS	1.29e+02	487	6	1.2	516	1	YBAF_SCHPO	HYPOTHETICAL 57.6 KD P	1.29e+02
415	6	1.2	449	1	LIPP_PIG	TRIACYLGLYCEROL LIPASE	1.29e+02	488	6	1.2	519	1	KPYR_CANRA	PYRUVATE KINASE, ISOXY	1.29e+02
416	6	1.2	451	1	OCT6_RAT	OCTAMER-BINDING TRANS	1.29e+02	489	6	1.2	519	1	ALAB_HUMAN	ALPHA-1B ADRENERGIC RE	1.29e+02
417	6	1.2	456	1	YD8_SCHPO	HYPOTHETICAL 52.4 KD P	1.29e+02	490	6	1.2	521	1	Y029_MYCTU	HYPOTHETICAL 56.0 KD P	1.29e+02
418	6	1.2	460	1	YX16_MYCTU	HYPOTHETICAL 51.3 KD P	1.29e+02	491	6	1.2	523	1	YHUS_ECOLI	HYPOTHETICAL 59.4 KD P	1.29e+02
419	6	1.2	460	1	SUOX_CHICK	SULFITE OXIDASE (EC.1.	1.29e+02	492	6	1.2	527	1	DPE2_HUMAN	ASA POLYMERASE ESSION	1.29e+02
420	6	1.2	461	1	VE2_HEV09	REGULATORY PROTEIN E2	1.29e+02	493	6	1.2	527	1	AK_YEAST	ASPARIKINASE (EC 2.7.	1.29e+02
421	6	1.2	463	1	ARLY_SACCO	ARGININOSUCCINATE LYAS	1.29e+02	494	6	1.2	528	1	YCD8_YEAST	HYPOTHETICAL 58.8 KD P	1.29e+02
422	6	1.2	463	1	ARLY_YEAST	ARGININOSUCCINATE LYAS	1.29e+02	495	6	1.2	530	1	PRO9_YEAST	PRE-MRNA SPLICING FCT	1.29e+02
423	6	1.2	465	1	LIPP_RABIT	TRIACYLGLYCEROL LIPASE	1.29e+02	496	6	1.2	531	1	FM01_PIG	DIMETHYLANILINE MONOOX	1.29e+02
424	6	1.2	466	1	CYP8_CAEEL	PEPTIDYL-PROXYL CIS-TR	1.29e+02	497	6	1.2	532	1	FM01_RAT	DIMETHYLANILINE MONOOX	1.29e+02
425	6	1.2	466	1	ALIA_RAT	ALPHA-1A ADRENERGIC RE	1.29e+02	498	6	1.2	532	1	FM01_MOUSE	DIMETHYLANILINE MONOOX	1.29e+02
426	6	1.2	466	1	ALIA_RABIT	ALPHA-1A ADRENERGIC RE	1.29e+02	499	6	1.2	534	1	YAJE_SCHPO	HYPOTHETICAL 59.0 KD P	1.29e+02
427	6	1.2	466	1	ALIA_HUMAN	ALPHA-1A ADRENERGIC RE	1.29e+02	500	6	1.2	534	1	FM01_RABIT	DIMETHYLANILINE MONOOX	1.29e+02
428	6	1.2	466	1	ALIA_BOVIN	ALPHA-1A ADRENERGIC RE	1.29e+02	501	6	1.2	542	1	RESB_BACSU	RESB PROTEIN	1.29e+02
429	6	1.2	466	1	ALIA_BOVIN	ALPHA-1A ADRENERGIC RE	1.29e+02	502	6	1.2	547	1	FIB1_ADE40	FIBER PROTEIN 1	1.29e+02
430	6	1.2	467	1	ATP1_CRIFA	ATP SYNTHASE BETA CHAI	1.29e+02	503	6	1.2	549	1	TREB_ECOLI	PROBABLE CYTOPLASMIC T	1.29e+02
431	6	1.2	468	1	NAH_SCHPO	NAT(+)/H(+) ANTIPORTER	1.29e+02	504	6	1.2	553	1	PPCK_CANAL	PHOSPHOENOLPYRUVATE CA	1.29e+02
432	6	1.2	469	1	ATPB_STRMU	ATP SYNTHASE BETA CHAI	1.29e+02	505	6	1.2	554	1	MHPA_ECOLI	3-(3-HYDROXY-PHENYL)P	1.29e+02
433	6	1.2	469	1	ATPB_BACFI	ATP SYNTHASE BETA CHAI	1.29e+02	506	6	1.2	554	1	NBL4_MOUSE	NBL4 PROTEIN	1.29e+02
434	6	1.2	470	1	YICO_ECOLI	HYPOTHETICAL 49.9 KD P	1.29e+02	507	6	1.2	557	1	YAH6_SCHPO	HYPOTHETICAL 62.7 KD P	1.29e+02
435	6	1.2	470	1	ALIA_ORYLA	ALPHA-1A ADRENERGIC RE	1.29e+02	508	6	1.2	558	1	DHE3_MOUSE	GLUTAMATE DEHYDROGENAS	1.29e+02
436	6	1.2	471	1	APPA_CAEEL	PUTATIVE ACID PHOSPHAT	1.29e+02	509	6	1.2	558	1	DHE3_HUMAN	GLUTAMATE DEHYDROGENAS	1.29e+02
437	6	1.2	472	1	APPA_BACME	ATP SYNTHASE BETA CHAI	1.29e+02	510	6	1.2	558	1	DHE3_RAT	GLUTAMATE DEHYDROGENAS	1.29e+02
438	6	1.2	474	1	NIFE_AZOVI	NITROGENASE IRON-MOLYB	1.29e+02	511	6	1.2	562	1	FIB1_ADE41	FIBER PROTEIN 1	1.29e+02
439	6	1.2	475	1	U2AF_HUMAN	SPLICING FACTOR U2AF 6	1.29e+02	512	6	1.2	564	1	HSP2_CHICK	HEAT SHOCK FACTOR PROT	1.29e+02
440	6	1.2	475	1	U2AF_MOUSE	SPLICING FACTOR U2AF 6	1.29e+02	513	6	1.2	566	1	HEMA_TAZIN	HEMAGGUTININ PRECURSO	1.29e+02
441	6	1.2	476	1	BM3B_MOUSE	BONE MORPHOGENETIC PRO	1.29e+02	514	6	1.2	566	1	MP12_HUMAN	M-PHASE INOUCER PROSH	1.29e+02
442	6	1.2	476	1	BM3B_MOUSE	BONE MORPHOGENETIC PRO	1.29e+02	515	6	1.2	566	1	HEMA_TAZIN	HEMAGGUTININ PRECURSO	1.29e+02
443	6	1.2	477	1	BM3B_MOUSE	BONE MORPHOGENETIC PRO	1.29e+02	516	6	1.2	566	1	HEMA_TAZIN	HEMAGGUTININ PRECURSO	1.29e+02
444	6	1.2	478	1	BM3B_MOUSE	BONE MORPHOGENETIC PRO	1.29e+02	517	6	1.2	566	1	HEMA_TAZIN	HEMAGGUTININ PRECURSO	1.29e+02
445	6	1.2	479	1	VGIC_HSV2G	GLYCOPROTEIN C PRECURS	1.29e+02	518	6	1.2	566	1	HEMA_TAZIN	HEMAGGUTININ PRECURSO	1.29e+02
446	6	1.2	479	1	ICER_BOMO	CASPASE-8 PRECURSOR	1.29e+02	519	6	1.2	568	1	GGL_MOUSE	GAMMA-GLUTAMYLTRANSPEP	1.29e+02
447	6	1.2	480	1	VGIC_HSV23	GLYCOPROTEIN C PRECURS	1.29e+02	520	6	1.2	568	1	GGL_MOUSE	GAMMA-GLUTAMYLTRANSPEP	1.29e+02
448	6	1.2	480	1	VGIC_HSV23	GLYCOPROTEIN C PRECURS	1.29e+02	521	6	1.2	568	1	GGL_MOUSE	GAMMA-GLUTAMYLTRANSPEP	1.29e+02
449	6	1.2	481	1	VGIC_HSV23	GLYCOPROTEIN C PRECURS	1.29e+02	522	6	1.2	569	1	YAR3_SCHPO	HYPOTHETICAL 61.1 KD P	1.29e+02
450	6	1.2	481	1	VGIC_HSV23	GLYCOPROTEIN C PRECURS	1.29e+02	523	6	1.2	569	1	YAR3_SCHPO	HYPOTHETICAL 61.1 KD P	1.29e+02
451	6	1.2	482	1	VGIC_HSV23	GLYCOPROTEIN C PRECURS	1.29e+02	524	6	1.2	570	1	SM4_CAEEL	MODULATION PROTEIN U (1.29e+02
452	6	1.2	484	1	NEP1_NAFRO	VIRULENCE-RELATED PROT	1.29e+02	525	6	1.2	570	1	SM4_CAEEL	MODULATION PROTEIN U (1.29e+02
453	6	1.2	484	1	NEP1_NAFRO	VIRULENCE-RELATED PROT	1.29e+02	526	6	1.2	570	1	SM4_CAEEL	MODULATION PROTEIN U (1.29e+02
454	6	1.2	490	1	CLK3_MOUSE	PROTEIN KINASE CLK3 (E	1.29e+02	527	6	1.2	574	1	KPYR_HUMAN	PYRUVATE KINASE, ISOXY	1.29e+02
455	6	1.2	490	1	CLK3_MOUSE	PROTEIN KINASE CLK3 (E	1.29e+02	528	6	1.2	574	1	KPYR_HUMAN	PYRUVATE KINASE, ISOXY	1.29e+02
456	6	1.2	490	1	CLK3_MOUSE	PROTEIN KINASE CLK3 (E	1.29e+02	529	6	1.2	574	1	KPYR_HUMAN	PYRUVATE KINASE, ISOXY	1.29e+02
457	6	1.2	492	1	CLK3_MOUSE	PROTEIN KINASE CLK3 (E	1.29e+02	530	6	1.2	574	1	KPYR_HUMAN	PYRUVATE KINASE, ISOXY	1.29e+02
458	6	1.2	492	1	CLK3_MOUSE	PROTEIN KINASE CLK3 (E	1.29e+02	531	6	1.2	578	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
459	6	1.2	494	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	532	6	1.2	583	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
460	6	1.2	496	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	533	6	1.2	583	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
461	6	1.2	498	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	534	6	1.2	590	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
462	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	535	6	1.2	590	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
463	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	536	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
464	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	537	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
465	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	538	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
466	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	539	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
467	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	540	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
468	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	541	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
469	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	542	6	1.2	592	1	KPYR_MOUSE	PUTATIVE SERINE CARBOX	1.29e+02
470	6	1.2	500	1	PHYS_AVESE	HYDROGENASE TRANSSCRIPT	1.29e+02	543	6	1.2	592				

243	6	1.2	269	1	COX3_TRTU	CYTOCHROME C OXIDASE P	1.29e+02	316	6	1.2	355	1	OPSG_CHICK	GREEN-SENSITIVE OPSIN	1.29e+02
244	6	1.2	271	1	Y4BG_RHTN	HYPOTHETICAL PROTEIN Y	1.29e+02	317	6	1.2	355	1	OPSB_GECE	BLUE-SENSITIVE OPSIN P	1.29e+02
245	6	1.2	271	1	ECBL_CANAL	CELL ELONGATION PROTEI	1.29e+02	318	6	1.2	355	1	OPSB_ANOCA	BLUE-SENSITIVE OPSIN	1.29e+02
246	6	1.2	274	1	XAP5_HUMAN	XAP-5 PROTEIN (FRAGMENT	1.29e+02	319	6	1.2	355	1	CMAT_STRE	CARMINOMYCIN 4-O-METH	1.29e+02
247	6	1.2	275	1	SRP1_SCHPO	SRP1 PROTEIN	1.29e+02	320	6	1.2	356	1	RS41_ARATH	ARGININE/SERINE-RICH S	1.29e+02
248	6	1.2	277	1	ICE3_CRITO	ADOPAIN PRECURSOR (EC	1.29e+02	321	6	1.2	359	1	OPSG_MOUSE	GREEN-SENSITIVE OPSIN	1.29e+02
249	6	1.2	277	1	ICE3_RAT	ADOPAIN PRECURSOR (EC	1.29e+02	322	6	1.2	361	1	OPSB_MOUSE	BLUE-SENSITIVE OPSIN	1.29e+02
250	6	1.2	277	1	ICE3_RAT	ADOPAIN PRECURSOR (EC	1.29e+02	323	6	1.2	361	1	OPSB_CHICK	BLUE-SENSITIVE OPSIN	1.29e+02
251	6	1.2	277	1	ICE3_HUMAN	ADOPAIN PRECURSOR (EC	1.29e+02	324	6	1.2	364	1	OPSB_CHICK	BLUE-SENSITIVE OPSIN	1.29e+02
252	6	1.2	277	1	ICE3_MOUSE	ADOPAIN PRECURSOR (EC	1.29e+02	325	6	1.2	364	1	OPSB_MOUSE	BLUE-SENSITIVE OPSIN	1.29e+02
253	6	1.2	281	1	TEFL_MOUSE	TRANSCRIPTION FACTOR E	1.29e+02	326	6	1.2	364	1	OPSG_HUMAN	GREEN-SENSITIVE OPSIN	1.29e+02
254	6	1.2	283	1	YB89_METUJ	HYPOTHETICAL PROTEIN M	1.29e+02	327	6	1.2	365	1	OPSG_MOUSE	GREEN-SENSITIVE OPSIN	1.29e+02
255	6	1.2	283	1	DAAA_BACSH	D-ALANINE AMINOTRANSFE	1.29e+02	328	6	1.2	365	1	YAGB_SCHPO	HYPOTHETICAL 4.2 KD P	1.29e+02
256	6	1.2	285	1	ATPG_BACME	ATP SYNTHASE GAMMA CHA	1.29e+02	329	6	1.2	366	1	DP3B_SALTY	DNA POLYMERASE III, BE	1.29e+02
257	6	1.2	285	1	VNSI_INCCA	3-CHLOROSENCOATE-3,4-D	1.29e+02	330	6	1.2	367	1	MTCL_CHV13	DNA POLYMERASE III, BE	1.29e+02
258	6	1.2	288	1	CBAB_ALCSP	NONSTRUCTURAL PROTEIN	1.29e+02	331	6	1.2	367	1	MTCL_CHV13	MODIFICATION METHYLASE	1.29e+02
259	6	1.2	288	1	MOVV_PSVJ	CELL-TO-CELL MOVEMENT	1.29e+02	332	6	1.2	368	1	DP50_BRABE	POU DOMAIN PROTEIN 2P-	1.29e+02
260	6	1.2	289	1	YFK2_YEAST	HYPOTHETICAL 31.9 KD P	1.29e+02	333	6	1.2	372	1	ALAT_CYPCA	ALPHA-1-ANTITRYPSIN HO	1.29e+02
261	6	1.2	289	1	MT61_METUJ	PROBABLE MODIFICATION	1.29e+02	334	6	1.2	372	1	LEML_MOUSE	L-SELECTIN PRECURSOR	1.29e+02
262	6	1.2	289	1	ATPG_KILUA	ATP SYNTHASE GAMMA CHA	1.29e+02	335	6	1.2	373	1	YF34_YEAST	HYPOTHETICAL 40.4 KD P	1.29e+02
263	6	1.2	290	1	SSO1_YEAST	SSO1 PROTEIN	1.29e+02	336	6	1.2	373	1	ICEB_MOUSE	CASPASE-11 PRECURSOR	1.29e+02
264	6	1.2	290	1	HTRU_ECOLI	HTRU PROTEIN	1.29e+02	337	6	1.2	374	1	VEF1_NPVOP	VERY LATE EXPRESSION F	1.29e+02
265	6	1.2	294	1	DRAG_RHORU	ADP-RIBOSYLGLYCERYLROL	1.29e+02	338	6	1.2	376	1	YREC_VIBCH	HYPOTHETICAL 41.3 KD P	1.29e+02
266	6	1.2	296	1	Z124_HUMAN	ZINC FINGER PROTEIN 12	1.29e+02	339	6	1.2	376	1	YB11_YEAST	HYPOTHETICAL 44.4 KD P	1.29e+02
267	6	1.2	296	1	PERX_BRARA	PEROXIDASE P7 (EC 1.11	1.29e+02	340	6	1.2	377	1	ICE4_HUMAN	CASPASE-4 PRECURSOR (E	1.29e+02
268	6	1.2	297	1	VGLG_HRSVJ	MAJOR SURFACE GLYCOPRO	1.29e+02	341	6	1.2	378	1	MOBL_THITE	MOBILIZATION PROTEIN M	1.29e+02
269	6	1.2	298	1	V727_METUJ	HYPOTHETICAL PROTEIN M	1.29e+02	342	6	1.2	381	1	GALI_ECOLI	GALACTOKINASE (EC 2.7.	1.29e+02
270	6	1.2	298	1	VGLG_HRSVJ	MAJOR SURFACE GLYCOPRO	1.29e+02	343	6	1.2	381	1	CYB_SARNA	CYCLOCHROME B (EC 1.10.	1.29e+02
271	6	1.2	298	1	VGLG_HRSVJ	MAJOR SURFACE GLYCOPRO	1.29e+02	344	6	1.2	381	1	THII_METUJ	PROBABLE THIAMIN BIOSY	1.29e+02
272	6	1.2	298	1	VEGB_HRSVA	FERROUS IRON TRANSPORT	1.29e+02	345	6	1.2	381	1	CYB_DASHA	CYCLOCHROME B (EC 1.10.	1.29e+02
273	6	1.2	301	1	VEGB_HRSVA	FERROUS IRON TRANSPORT	1.29e+02	346	6	1.2	382	1	NCAP_CVPPU	NUCLEOCAPSID PROTEIN	1.29e+02
274	6	1.2	301	1	MCP_BPF41	MAJOR CAPSID PROTEIN	1.29e+02	347	6	1.2	382	1	NCAP_CVPPU	NUCLEOCAPSID PROTEIN	1.29e+02
275	6	1.2	304	1	ICE7_MOUSE	CASPASE-7 PRECURSOR (E	1.29e+02	348	6	1.2	382	1	NCAP_CVPP8	NUCLEOCAPSID PROTEIN	1.29e+02
276	6	1.2	306	1	THTR_YEAST	PUTATIVE THIOSULFATE S	1.29e+02	349	6	1.2	382	1	NCAP_CVPP8	NUCLEOCAPSID PROTEIN	1.29e+02
277	6	1.2	306	1	KDGR_ERCHC	PROTEIN DEGRADATION REP	1.29e+02	350	6	1.2	382	1	ICEB_XENLA	INTERLEUKIN-1 BETA CON	1.29e+02
278	6	1.2	309	1	WAPP_COCIM	WALL-ASSOCIATED PROTEI	1.29e+02	351	6	1.2	382	1	CYB_MONDO	CYCLOCHROME B (EC 1.10.	1.29e+02
279	6	1.2	311	1	CYP9_CAEEL	PEPTIDYL-PROLYL CIS-TR	1.29e+02	352	6	1.2	382	1	NCAP_CVPPM	NUCLEOCAPSID PROTEIN	1.29e+02
280	6	1.2	312	1	SUFI_HAEIN	SUFI PROTEIN HOMOLOG P	1.29e+02	353	6	1.2	383	1	VE2_HPV15	REGULATORY PROTEIN E2.	1.29e+02
281	6	1.2	312	1	ATO_DROME	ATONAL PROTEIN	1.29e+02	354	6	1.2	383	1	GYRA_MYCGA	DNA GYRASE SUBUNIT A	1.29e+02
282	6	1.2	312	1	PDXK_HUMAN	PYRIDOXINE KINASE (EC	1.29e+02	355	6	1.2	385	1	KUPI_CAEEL	KUP-1 PROTEIN	1.29e+02
283	6	1.2	314	1	NIKB_ECOLI	NICKEL TRANSPORT STYE	1.29e+02	356	6	1.2	385	1	INTC_ECOLI	PUTATIVE PROPHAGE SF6-	1.29e+02
284	6	1.2	314	1	OLFI_HUMAN	OLFACTORY RECEPTOR-LIK	1.29e+02	357	6	1.2	386	1	SUCC_RICPR	SUCCINYL-COA SYNTHETAS	1.29e+02
285	6	1.2	319	1	COA2_POMV3	COAT PROTEIN VP2 (CONT	1.29e+02	358	6	1.2	386	1	ICEA_XENLA	INTERLEUKIN-1 BETA CON	1.29e+02
286	6	1.2	319	1	COA2_POMV3	COAT PROTEIN VP2 (CONT	1.29e+02	359	6	1.2	388	1	OXYR_RAT	OXYTOCIN RECEPTOR (OT-	1.29e+02
287	6	1.2	323	1	YECP_ECOLI	COAT PROTEIN VP2 (CONT	1.29e+02	360	6	1.2	388	1	OXYR_MOUSE	OXYTOCIN RECEPTOR (OT-	1.29e+02
288	6	1.2	323	1	YECP_ECOLI	COAT PROTEIN VP2 (CONT	1.29e+02	361	6	1.2	388	1	GSPL_ECOLI	PROBABLE GENERAL SECRE	1.29e+02
289	6	1.2	323	1	ICEB_DROME	CASPASE-1 PRECURSOR (E	1.29e+02	362	6	1.2	389	1	Y107_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02
290	6	1.2	327	1	WZB2_ECOLI	CHAIN LENGTH DETERMINA	1.29e+02	363	6	1.2	393	1	PURT_METUJ	PROBABLE PHOSPHORIBOSY	1.29e+02
291	6	1.2	328	1	NIR3_AZOBR	NIFER3-LIKE PROTEIN.	1.29e+02	364	6	1.2	394	1	FTS2_AZOVI	CELL DIVISION PROTEIN	1.29e+02
292	6	1.2	334	1	API_RAT	TRANSCRIPTION FACTOR A	1.29e+02	365	6	1.2	394	1	FTS2_PSENE	CELL DIVISION PROTEIN	1.29e+02
293	6	1.2	334	1	API_MOUSE	TRANSCRIPTION FACTOR A	1.29e+02	366	6	1.2	395	1	YHAC_ECOLI	HYPOTHETICAL 45.2 KD P	1.29e+02
294	6	1.2	336	1	YNG6_YEAST	CASPASE PRECURSOR (EC	1.29e+02	367	6	1.2	395	1	NODC_AZOCA	N-ACETYLGUCOSAMINYLTR	1.29e+02
295	6	1.2	339	1	ICE_DROME	CASPASE PRECURSOR (EC	1.29e+02	368	6	1.2	395	1	ILV5_YEAST	KEUOL-ACID REDUCTOISOM	1.29e+02
296	6	1.2	340	1	MBR1_YEAST	MBR1 PROTEIN.	1.29e+02	369	6	1.2	397	1	TYRB_ECOLI	AROMATIC-AMINO-ACID AM	1.29e+02
297	6	1.2	341	1	PLDB_ECOLI	LYSOPHOSPHOLIPASE L2	1.29e+02	370	6	1.2	399	1	BR33_HUMAN	BOMBSA-1 RECEPTOR SUBT	1.29e+02
298	6	1.2	343	1	TAZR_RAT	THROMBOXANE A2 RECEPT	1.29e+02	371	6	1.2	399	1	COAT_MOUSE	COAT PROTEIN PRECURSOR	1.29e+02
299	6	1.2	343	1	HRCA_BACSU	HEAT-INHIBITABLE TRANSCE	1.29e+02	372	6	1.2	400	1	FTS2_PSEBU	CELL DIVISION PROTEIN	1.29e+02
300	6	1.2	344	1	ILVC_LACLA	KETOL-ACID REDUCTOISOM	1.29e+02	373	6	1.2	401	1	YNZ4_YEAST	HYPOTHETICAL 46.2 KD P	1.29e+02
301	6	1.2	344	1	STST_RAUSE	STRIKOSTIDINE SYNTHASE	1.29e+02	374	6	1.2	401	1	RU3_TOXCA	60S RIBOSOMAL PROTEIN	1.29e+02
302	6	1.2	344	1	HEMA_IAMAR	HEMAGGLUTININ PRECURSO	1.29e+02	375	6	1.2	402	1	ILBC_RAT	INTERLEUKIN-1 BETA CON	1.29e+02
303	6	1.2	344	1	HEMA_IAXIA	HEMAGGLUTININ PRECURSO	1.29e+02	376	6	1.2	402	1	ILBC_MOUSE	INTERLEUKIN-1 BETA CON	1.29e+02
304	6	1.2	346	1	HEMA_IATAI	HEMAGGLUTININ PRECURSO	1.29e+02	377	6	1.2	403	1	HMPA_ALCEU	FLAVOENOPROTEIN (HAEH	1.29e+02
305	6	1.2	346	1	Y131_BACSU	HYPOTHETICAL 37.0 KD P	1.29e+02	378	6	1.2	404	1	GLUC_MCTU	GLUCOSE-1-PHOSPHATE AD	1.29e+02
306	6	1.2	347	1	Y131_HAEIN	HYPOTHETICAL PROTEIN H	1.29e+02	379	6	1.2	406	1	RTA_RAPPH	ISOCITRATE DEHYDROGENA	1.29e+02
307	6	1.2	347	1	UTR2_YEAST	VIOLET-SENSITIVE OPSIN	1.29e+02	380	6	1.2	407	1	IDH_SPHYA	PUTATIVE AMINOTRANSFER	1.29e+02
308	6	1.2	349	1	UTR2_YEAST	UTR2 PROTEIN (UNKNOWN	1.29e+02	381	6	1.2	409	1	YNP2_YEAST	HYPOTHETICAL 46.2 KD P	1.29e+02
309	6	1.2	350	1	CGM6_HUMAN	CARCINOEMBRYONIC ANTIG	1.29e+02	382	6	1.2	411	1	MSK1_MEDSA	GLYCOCEN SYNTHASE KINA	1.29e+02
310	6	1.2	352	1	MUTY_ECOLI	A/G-SPECIFIC ADENINE G	1.29e+02	383	6	1.2	413	1	EFIA_METH	ELONGATION FACTOR 1-AL	1.29e+02
311	6	1.2	352	1	WNT2_DROME	PROTEIN DMMT-2 PRECURS	1.29e+02	384	6	1.2	415	1	CFRR_MOUSE	CORTICOTROPIN RELEASEIN	1.29e+02
312	6	1.2	352	1	OPSD_ANGAN	RHODOPSIN, DEEP-SEA FO	1.29e+02	385	6	1.2	415	1	ERK1_DICDI	EXTRACELLULAR SIGNAL-R	1.29e+02
313	6	1.2	352	1	OPSD_ANGAN	RHODOPSIN, DEEP-SEA FO	1.29e+02	386	6	1.2	415	1	CFRR_RAT	CORTICOTROPIN RELEASEIN	1.29e+02
314	6	1.2	354	1	HEM6_HUMAN	HEM6_HUMAN	1.29e+02	387	6	1.2	415	1	CFRR_SHEEP	CORTICOTROPIN RELEASEIN	1.29e+02
315	6	1.2	354	1	HEM6_MOUSE	COPROPHOPHYRINNGEN III	1.29e+02	388	6	1.2	416	1	ICE9_HUMAN	CASPASE-9 PRECURSOR (E	1.29e+02

97	1.4	488	1	VE2_HPV49	REGULATORY PROTEIN E2.	2.20e+00	170	1.2	161	1	SECE_MYCTU	PROBABLE PREPROTEIN TR	1.29e+02
98	1.4	493	1	VE2_HPV19	REGULATORY PROTEIN E2.	2.20e+00	171	1.2	161	1	ATPX_HOBU	ATP SYNTHASE B' CHAIN	1.29e+02
99	1.4	494	1	SEF4_HUMAN	SPLICING FACTOR, ARGIN	2.20e+00	172	1.2	162	1	19KD_MYCT	19 KD LIPOPROTEIN ANTI	1.29e+02
100	1.4	497	1	VE2_HPV20	REGULATORY PROTEIN E2.	2.20e+00	173	1.2	170	1	PLAS_LYCS	PLASTOCYANIN PRECURSOR	1.29e+02
101	1.4	498	1	VE2_HPV8	REGULATORY PROTEIN E2.	2.20e+00	174	1.2	170	1	HEMX_PROMI	PURATIVE UROPORPHYRIN-	1.29e+02
102	1.4	502	1	VE2_HPV25	REGULATORY PROTEIN E2.	2.20e+00	175	1.2	175	1	HEMX_MOUSE	TRANSCRIPTION FACTOR H	1.29e+02
103	1.4	503	1	VE2_HPV21	REGULATORY PROTEIN E2.	2.20e+00	176	1.2	175	1	HES3_RAT	TRANSCRIPTION FACTOR H	1.29e+02
104	1.4	514	1	VE2_HPV5	REGULATORY PROTEIN E2.	2.20e+00	177	1.2	177	1	ATPD_CYACA	ATP SYNTHASE DELTA CHA	1.29e+02
105	1.4	514	1	VE2_HPV5	REGULATORY PROTEIN E2.	2.20e+00	178	1.2	178	1	YQAC_BACSU	HYPOTHETICAL 20.7 KD P	1.29e+02
106	1.4	516	1	CP23_HOBU	SERINE CARBOXYPEPTIDAS	2.20e+00	179	1.2	180	1	TRSE_DROER	FEMALE-SPECIFIC TRANSF	1.29e+02
107	1.4	615	1	IBR_HUMAN	LAMIN B RECEPTOR (INTE	2.20e+00	180	1.2	180	1	Y271_METJA	HYPOTHETICAL 21.0 KD P	1.29e+02
108	1.4	620	1	Y870_METJA	HYPOTHETICAL PROTEIN M	2.20e+00	181	1.2	181	1	CHMU_ERHME	MONOFUNCTIONAL CHORISM	1.29e+02
109	1.4	637	1	IBR_CHICK	LAMIN B RECEPTOR	2.20e+00	182	1.2	181	1	YHOL_YEAS	HYPOTHETICAL 21.0 KD P	1.29e+02
110	1.4	646	1	SR72_YEAS	SIGNAL RECOGNITION PAR	2.20e+00	183	1.2	184	1	TRSE_DROST	FEMALE-SPECIFIC TRANSF	1.29e+02
111	1.4	692	1	YK06_YEAS	HYPOTHETICAL 75.5 KD P	2.20e+00	184	1.2	185	1	Y073_BORBU	TRANSCRIPTION ANTIEMEM	1.29e+02
112	1.4	715	1	KM2_HUMAN	INTERFERON-REGULATED R	2.20e+00	185	1.2	185	1	NUSG_TREPA	HYPOTHETICAL PROTEIN B	1.29e+02
113	1.4	739	1	DD13_CAEEL	PURATIVE PRE-MRNA SPLI	2.20e+00	186	1.2	187	1	Y271_METJA	60S RIBOSOMAL PROTEIN	1.29e+02
114	1.4	759	1	YEN1_YEAS	HYPOTHETICAL 87.4 KD P	2.20e+00	187	1.2	187	1	RI18_RAT	60S RIBOSOMAL PROTEIN	1.29e+02
115	1.4	769	1	CIFK_MOUSE	VOLTAGE-GATED POTASSIU	2.20e+00	188	1.2	187	1	PDDB_ECOLI	PREPILIN PEPTIDASE DEP	1.29e+02
116	1.4	814	1	SLAP_BACAN	S-LAYER PROTEIN PRECUR	2.20e+00	189	1.2	195	1	MTLR_ECOLI	MANNITOL OPPERON REPRS	1.29e+02
117	1.4	901	1	SOK1_YEAS	SOKI PROTEIN	2.20e+00	190	1.2	195	1	F0MB_METJA	PURATIVE FUMARATE HYDR	1.29e+02
118	1.4	951	1	SFR8_HUMAN	SPLICING FACTOR, GAMMA	2.20e+00	191	1.2	196	1	UTDR_ECOLI	UTD OPERON REPRESSOR (1.29e+02
119	1.4	1070	1	PVDG_PLAKN	DUFFY RECEPTOR	2.20e+00	192	1.2	196	1	SFR2_CAEEL	PURATIVE SPLICING FACT	1.29e+02
120	1.4	1150	1	IRRI_YEAS	PURATIVE PRE-MRNA SPLI	2.20e+00	193	1.2	197	1	TRSE_DROME	FEMALE-SPECIFIC TRANSF	1.29e+02
121	1.4	1200	1	DDX8_CAEEL	PROBABLE ATP-DEPENDENT	2.20e+00	194	1.2	199	1	WC07_ADEMI	MAJOR CORE PROTEIN PRE	1.29e+02
122	1.4	1220	1	SLA1_HUMAN	CYTOSKELETON ASSEMBLY	2.20e+00	195	1.2	200	1	PUAC_STRLP	PUROMYCIN N-ACETYLTRAN	1.29e+02
123	1.4	1244	1	SLA1_YEAS	ERBB-3 RECEPTOR PROTEI	2.20e+00	196	1.2	203	1	YG3R_YEAS	CHLOROPLAST 30S RIBOSO	1.29e+02
124	1.4	1339	1	ERB3_RAT	ERBB-3 RECEPTOR PROTEI	2.20e+00	197	1.2	203	1	RG3R_YEAS	HYPOTHETICAL 22.2 KD P	1.29e+02
125	1.4	1342	1	ERB3_HUMAN	ERBB-3 RECEPTOR PROTEI	2.20e+00	198	1.2	205	1	GARI_YEAS	HYPOTHETICAL 24.0 KD P	1.29e+02
126	1.4	1466	1	NKCR_MOUSE	NK-TUMOR RECOGNITION P	2.20e+00	199	1.2	208	1	YXK2_CAEEL	GARI PROTEIN	1.29e+02
127	1.4	1523	1	SON_HUMAN	SON PROTEIN (SON3)	2.20e+00	200	1.2	209	1	Y113_YEAS	PROBABLE PROTEIN-TYROS	1.29e+02
128	1.4	2278	1	FBI1_YEAS	PROBABLE PHOSPHATIDYL	2.20e+00	201	1.2	209	1	FTSJ_ECOLI	CELL DIVISION PROTEIN	1.29e+02
129	1.2	35	1	PHI1_MYCTA	SPERM-SPECIFIC PROTEIN	1.29e+02	202	1.2	212	1	ATP6_TROHI	ATP SYNTHASE A CHAIN (1.29e+02
130	1.2	37	1	PR73_SCYCA	PROTAMINE Z3 (SCYLLATOR	1.29e+02	203	1.2	214	1	BT33_HUMAN	TRANSCRIPTION FACTOR B	1.29e+02
131	1.2	54	1	BVCP_NPVAC	DNA-BINDING PROTEIN (A	1.29e+02	204	1.2	214	1	RADC_HOCHA	DNA REPAIR PROTEIN RAD	1.29e+02
132	1.2	64	1	BVCP_NPVAC	DNA-BINDING PROTEIN (A	1.29e+02	205	1.2	216	1	OLID_HUMAN	OLFACTORY RECEPTOR-LIK	1.29e+02
133	1.2	68	1	YHVH_SALTU	HYPOTHETICAL PROTEIN (A	1.29e+02	206	1.2	216	1	VIF_HVZD2	OLFACTORY RECEPTOR-LIK	1.29e+02
134	1.2	68	1	HSP1_TACAC	SPERM PROTEIN	1.29e+02	207	1.2	216	1	OLIA_HUMAN	VIROIN INFECTIVITY FAC	1.29e+02
135	1.2	75	1	YKGI_HAUSA	GAS VESICLE PROTEIN A,	1.29e+02	208	1.2	216	1	LUXM_VIBHA	LUXM PROTEIN	1.29e+02
136	1.2	83	1	YKGI_ECOLI	HYPOTHETICAL 9.1 KD PR	1.29e+02	209	1.2	219	1	REP2_SCHPO	TRANSCRIPTIONAL ACTIVA	1.29e+02
137	1.2	85	1	RPOD_HYCAN	DNA-DIRECTED RNA POLYM	1.29e+02	210	1.2	222	1	PHZ6_PSEFL	PHENAZINE BIOSYNTHESIS	1.29e+02
138	1.2	87	1	MATB_HUMAN	CHLORIDE CONDUCTANCE I	1.29e+02	211	1.2	225	1	KAD2_YEAS	ADENYLATE KINASE 2 (EC	1.29e+02
139	1.2	91	1	YAAD_SCHPO	HYPOTHETICAL 10.5 KD P	1.29e+02	212	1.2	225	1	ATP6_PONPP	ATP SYNTHASE A CHAIN (1.29e+02
140	1.2	91	1	PHI1_MYCTD	SPERM-SPECIFIC PROTEIN	1.29e+02	213	1.2	227	1	NODM_BRAJA	MODULATION PROTEIN W.	1.29e+02
141	1.2	95	1	ACP_STRRM	OXYTETRACYCLINE POLYME	1.29e+02	214	1.2	229	1	CMF3_BACSU	CONF OPERON PROTEIN 3.	1.29e+02
142	1.2	99	1	PLAS_CAPBU	PLASTOCYANIN	1.29e+02	215	1.2	229	1	S3AC_BACSU	STAGE-III SPOREATION	1.29e+02
143	1.2	99	1	PLAS_CAPBU	PLASTOCYANIN	1.29e+02	216	1.2	229	1	VRP_LAMBD	REPLICATION PROTEIN P.	1.29e+02
144	1.2	99	1	IT12_SINAI	TRYPSIN INHIBITOR 2 PR	1.29e+02	217	1.2	233	1	SCR3_MESAU	SYNAPTONEMAL COMPLEX P.	1.29e+02
145	1.2	100	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	218	1.2	234	1	ATP3_MESAU	ATP SYNTHASE DELTA CHA	1.29e+02
146	1.2	100	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	219	1.2	237	1	DEOD_HAEIN	PURINE NUCLEOSIDE PHOS	1.29e+02
147	1.2	103	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	220	1.2	238	1	YJFH_ECOLI	HYPOTHETICAL TNA/RNA	1.29e+02
148	1.2	104	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	221	1.2	244	1	TPIS_MYCPN	TRIOSEPHOSPHATE ISOMER	1.29e+02
149	1.2	106	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	222	1.2	246	1	DMG_SOUNC	PROTEOLIPID PROTEIN DM	1.29e+02
150	1.2	110	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	223	1.2	246	1	SDGF_MOUSE	SCHWANNOMA-DERIVED GRO	1.29e+02
151	1.2	114	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	224	1.2	253	1	CYSZ_ECOLI	4-1BB LIGAND (4-1BBL)	1.29e+02
152	1.2	117	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	225	1.2	254	1	ADBL_HUMAN	ALCOHOL DEHYDROGENASE	1.29e+02
153	1.2	118	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	226	1.2	254	1	ADBL_HUMAN	ALCOHOL DEHYDROGENASE	1.29e+02
154	1.2	120	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	227	1.2	255	1	ATP6_SCHPO	ATP SYNTHASE A CHAIN P	1.29e+02
155	1.2	120	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	228	1.2	257	1	ATP6_SCHPO	ATP SYNTHASE A CHAIN P	1.29e+02
156	1.2	122	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	229	1.2	260	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
157	1.2	125	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	230	1.2	260	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
158	1.2	131	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	231	1.2	261	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
159	1.2	133	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	232	1.2	261	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
160	1.2	135	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	233	1.2	262	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
161	1.2	138	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	234	1.2	262	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
162	1.2	139	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	235	1.2	263	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
163	1.2	142	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	236	1.2	264	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
164	1.2	142	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	237	1.2	265	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
165	1.2	146	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	238	1.2	266	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
166	1.2	154	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	239	1.2	266	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
167	1.2	155	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	240	1.2	268	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
168	1.2	157	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	241	1.2	269	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02
169	1.2	158	1	YNTB_YEAS	VERY HYPOTHETICAL 11.8	1.29e+02	242	1.2	269	1	YNTB_YEAS	ATP SYNTHASE A CHAIN P	1.29e+02

Query Match 1.4%; Score 7; DB 2; Length 58;
 Best Local Similarity 100.0%; Pred. No. 4.44e+00;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RSRSRSR 13
 |||||
 QY .150 RSRSRSR 156

RESULT 15
 ENTRY S39425 #type complete
 TITLE protamine P1 - duckbill platypus
 ORGANISM #formal_name Ornithorhynchus anatinus #common_name duckbill
 platypus
 DATE 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change
 29-Jan-1999

ACCESSIONS S39425
 REFERENCE S39424
 #authors Retief, J.D.; Winkfein, R.J.; Dixon, G.H.
 #journal Eur. J. Biochem. (1993) 218:457-461
 #title Evolution of the monotremes. The sequences of the protamine
 P1 genes of platypus and echidna.
 #cross-references MIMD:94094837
 #accession S39425
 #status preliminary
 #molecule_type DNA
 #residues 1-61 #label RET
 ##cross-references EMBL:226849

GENETICS 47/1
 #introns
 CLASSIFICATION #superfamily sperm histone
 SUMMARY #length 61 #molecular_weight 7871 #checksum 4490

Query Match 1.4%; Score 7; DB 2; Length 61;
 Best Local Similarity 100.0%; Pred. No. 4.44e+00;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 RSRSRSR 12
 |||||
 QY 150 RSRSRSR 156

Search completed: Tue Aug 10 11:43:48 1999
 Job time : 94 secs.

OY 130 RDSVSSVS 137

RESULT 10
ENTRY 138728 #type complete
TITLE epidermal growth factor receptor kinase substrate - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Jun-1998

ACCESSIONS 138728
REFERENCE 138728

#authors Wong, W.T.; Carlomagno, F.; Druck, T.; Barletta, C.; Croce, C.M.; Huebner, K.; Kraus, M.H.; Di Fiore, P.P.
#journal Oncogene (1994) 9:3057-3061
#title Evolutionary conservation of the ERS8 gene and its mapping to human chromosome 12q23-q24.
#cross-references MIM:94366758
#accession 138728
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-822 #label RES
#cross-references EMBL:U12535; NID:9530822; PID:9530823

GENETICS ERS8
#gene length 822 #molecular-weight 91881 #checksum 5343

SUMMARY
Query Match 1.6%; Score 8; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 RDSVSSVS 60
|||||||
OY 130 RDSVSSVS 137

RESULT 11
ENTRY S21080 #type fragment
TITLE sperm protein EMI - minor jackknife clam (fragment)
ORGANISM #formal_name EMI - minor #common_name minor jackknife clam
DATE 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 23-Feb-1997

ACCESSIONS S21080
REFERENCE S21080

#authors Giannocotti, V.; Buratti, E.; Santucci, A.; Neri, P.; Crane-Robinson, C.
#journal Biochim. Biophys. Acta (1992) 1119:296-302
#title Molluscan sperm proteins: Emsis minor.
#cross-references MIM:92190249
#accession S21080
#molecule_type protein
#residues 1-34 #label G1A
KEYWORDS DNA binding; nucleus
SUMMARY length 34 #checksum 6188

Query Match 1.4%; Score 7; DB 2; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.44e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 RSRRSR 12
|||||||
OY 150 RSRRSR 156

RESULT 12
ENTRY C58213 #type complete
TITLE protamine II American alligator
ORGANISM #formal_name Alligator mississippiensis #common_name American alligator
DATE 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 09-May-1997

ACCESSIONS C58213
REFERENCE A58208

#authors Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice,

P.; Bell, J.E.; Sharp, D.J.; Kiss, A.J.; Hunt, D.F.; Arnold, D.P.; Russ, M.M.; Shabanowitz, J.; Ausio, J.
J. Biol. Chem. (1996) 271:23547-23557
#title Protamines of reptiles.
#accession C58213
#status preliminary
#molecule_type protein
#residues 1-56 #label HUN
CLASSIFICATION #superfamily sperm histone
SUMMARY length 56 #molecular-weight 7537 #checksum 6509

Query Match 1.4%; Score 7; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 4.44e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RSRRSR 13
|||||||
OY 150 RSRRSR 156

RESULT 13
ENTRY S34045 #type complete
TITLE protamine - North American opossum
ORGANISM #formal_name Didelphis virginiana, Didelphis marsupialis
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Sep-1997

ACCESSIONS S34045
REFERENCE S34045

#authors Winkler, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.
#journal Eur. J. Biochem. (1993) 215:63-72
#title Characterization of a marsupial sperm protamine gene and its transcripts from the North American opossum (Didelphis marsupialis).
#cross-references MIM:93345500
#accession S34045
#status preliminary
#molecule_type DNA
#residues 1-58 #label WIN
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus
SUMMARY length 58 #molecular-weight 7941 #checksum 3630

Query Match 1.4%; Score 7; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 4.44e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RSRRSR 13
|||||||
OY 150 RSRRSR 156

RESULT 14
ENTRY A58208 #type complete
TITLE protamine I-1 painted turtle
ORGANISM #formal_name Chrysemys picta #common_name painted turtle
DATE 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 09-May-1997

ACCESSIONS A58208
REFERENCE A58208

#authors Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.J.; Kiss, A.J.; Hunt, D.F.; Arnold, D.P.; Russ, M.M.; Shabanowitz, J.; Ausio, J.
J. Biol. Chem. (1996) 271:23547-23557
#title Protamines of reptiles.
#accession A58208
#status preliminary
#molecule_type protein
#residues 1-58 #label HUN
CLASSIFICATION #superfamily sperm histone
SUMMARY length 58 #molecular-weight 7620 #checksum 2377

A.M.: Presseau, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Ray, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scifone, F.; Sekiuchi, J.;
Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taccini, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Yamakoshi, A.; Tanaka, T.;
Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassart, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
#cross-references EMBL:98044033
#accession A70039
#status preliminary; nucleic acid sequence not shown;
#molecule-type DNA
#residues 1-301 #label KUN
#cross-references GB:Z99121; GB:AL009126; NID:92635827; PID:e1186097;
#experimental_source strain 168

GENETICS
#gene yvfr
#superfamily ATP-binding cassette homology
#keywords P-loop
#feature 21-206
#domain ATP-binding cassette homology #label ABC
#region nucleotide-binding motif A (P-loop)
#length 301 #molecular-weight 33779 #checksum 1714

SUMMARY
Query Match 1.6%; Score 8; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 91 LKVEILE 98
11111111
QY 21 LKVEILE 28

RESULT 7
ENTRY S75895 #type complete
TITLE hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
ORGANISM #formal_name *Synechocystis* sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998

ACCESSIONS S75895
REFERENCE S74332
#authors Kaneo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.

#cross-references EMBL:97061201
#accession S75895
#status preliminary
#molecule-type DNA
#residues 1-495 #label KAN
#cross-references EMBL:D90913; GB:AB001339; NID:g1653348; PID:d1019087;
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996

CLASSIFICATION #superfamily xylulokinase

SUMMARY #length 495 #molecular-weight 54351 #checksum 4379

Query Match 1.6%; Score 8; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 338 LARSVDN 345
11111111
QY 91 LARSVDN 98

RESULT 8
ENTRY W2M147 #type complete
TITLE E2 protein - human papillomavirus type 47
ORGANISM #formal_name human papillomavirus type 47
#note host Homo sapiens (man)
DATE 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change
05-Sep-1997

ACCESSIONS D35324
REFERENCE A53324
#authors Kiyono, T.; Adachi, A.; Ishibashi, M.
#journal Virology (1990) 177:401-405
#title Genome organization and taxonomic position of human
papillomavirus type 47 inferred from its DNA sequence.
#cross-references EMBL:90281611
#accession D35324
#status translation not shown
#molecule-type DNA
#residues 1-506 #label KIT
#cross-references GB:M3205; NID:9333062; PID:9333067

CLASSIFICATION #superfamily papillomavirus E2 protein
#keywords DNA binding; early protein; transcription regulation
#length 506 #molecular-weight 57478 #checksum 9705

SUMMARY
Query Match 1.6%; Score 8; DB 1; Length 506;
Best Local Similarity 100.0%; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 302 RARSRS 309
11111111
QY 148 RARSRS 155

RESULT 9
ENTRY S39983 #type complete
TITLE eps8 protein - mouse
ORGANISM #formal_name *Mus musculus* #common_name house mouse
#variety 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
17-Mar-1999

ACCESSIONS S39983
REFERENCE S39983
#authors Fazioli, F.; Minichiello, L.; Matoska, V.; Castagnino, P.;
Miki, T.; Wong, W.T.; di Fiore, P.P.
#journal EMBO J. (1993) 12:3799-3808
#title Eps8, a substrate for the epidermal growth factor receptor
kinase, enhances EGF-dependent mitogenic signals.

#cross-references EMBL:121671; NID:g309216; PID:g309217
#accession S39983
#status preliminary
#molecule-type mRNA
#residues 1-821 #label FAZ
#cross-references EMBL:121671; NID:g309216; PID:g309217

CLASSIFICATION #superfamily SH3 homology
#feature 537-584
#domain SH3 homology #label SH3
#length 821 #molecular-weight 91737 #checksum 9175

SUMMARY
Query Match 1.6%; Score 8; DB 2; Length 821;
Best Local Similarity 100.0%; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 RSVSSVS 60
11111111

```

#accession B40082
#molecule_type DNA
#residues 44-159 ##label L12
#accession C40082
#molecule_type protein
#residues 6-15;23-59;93-103;148-170;174-178 ##label L13
REFERENCE
#authors Lin, L.F.H.; Armes, L.G.; Sommer, A.; Smith, D.J.; Collins, F.
#journal J. Biol. Chem. (1990) 265:8942-8947
#title Isolation and characterization of ciliary neurotrophic factor from rabbit sciatic nerves.
#cross-references MUID:90256829
#accession A36286
#molecule_type protein
#residues 6-15;22-59;93-103;140-147;148-169;174-178 ##label L14
COMMENT
  CNTF has the characteristics of a nonsecreted, cytoplasmic protein; however, a receptor for this protein is found exclusively in nervous and skeletal muscle tissues (see ciliary neurotrophic factor receptor).
  CNTF promotes survival, neurotransmitter synthesis, and neurite outgrowth in certain neuronal populations.
  #superfamily ciliary neurotrophic factor
  #growth factor; peripheral neuron
  #length 199 #molecular-weight 22662 #checksum 8387
SUMMARY
  Query Match 1.6%; Score 8; DB 1; Length 199;
  Best Local Similarity 100.0%; Pred. No. 5,18e-02;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 50 LDSVDGVP 57
  |||||
  QY 370 LDSVDGVP 377

RESULT 4 UNRTCF #type complete
ENTRY ciliary neurotrophic factor - rat
TITLE CNTF: survival factor
ALTERNATE_NAMES #formal_name Rattus norvegicus #common_name Norway rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 29-May-1998
ACCESSIONS S08144; A43007
REFERENCE S08144
#authors Stoeckli, K.A.; Lottspeich, F.; Sendtner, M.; Masikowski, P.; Carroll, P.; Goetz, R.; Lindholm, D.; Thoenen, H.
#journal Nature (1989) 342:920-923
#title Molecular cloning, expression and regional distribution of rat ciliary neurotrophic factor.
#cross-references MUID:90081871
#accession S08144
#molecule_type mRNA
#residues 1-200 ##label STL
#cross-references EMBL:X17457; NID:g55968; PID:g55969
#accession A43007
#molecule_type protein
#residues 20-25;29-40;47-60;65-92;113-154;161-166;184-200 ##label ST2
COMMENT
  CNTF has the characteristics of a nonsecreted, cytoplasmic protein; however, a receptor for this protein is found exclusively in nervous and skeletal muscle tissues (see ciliary neurotrophic factor receptor, PIR:I58141).
  CNTF promotes survival, neurotransmitter synthesis, and neurite outgrowth in certain neuronal populations.
  #superfamily ciliary neurotrophic factor
  #growth factor; peripheral neuron
  #length 200 #molecular-weight 22854 #checksum 2308
SUMMARY
  Query Match 1.6%; Score 8; DB 1; Length 200;
  Best Local Similarity 100.0%; Pred. No. 5,18e-02;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 50 LDSVDGVP 57

```

```

QY 370 LDSVDGVP 377
  |||||
  RESULT 5
  ENTRY J06125 #type complete
  TITLE U2 small nuclear ribonucleoprotein auxiliary factor small chain - fruit fly (Drosophila melanogaster)
  ORGANISM #formal_name Drosophila melanogaster
  DATE 23-Mar-1997 #sequence_revision 09-May-1997 #text_change 09-Sep-1997
  ACCESSIONS J06125
  REFERENCE J06125
  #authors Rudner, D.Z.; Kanaar, R.; Breger, K.S.; Rio, D.C.
  #journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:10333-10337
  #title Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects.
  #cross-references MUID:96413646
  #accession J06125
  #status nucleic acid sequence not shown
  ##molecule_type mRNA
  #residues 1-264 ##label RUD
  #cross-references GB:D67066; NID:g1621614; PID:g1621615
  #experimental_source embryo
  COMMENT
    This protein plays a role in 3' splice site selection. It binds between the branch site and the 3' splice site of the pre-mRNA and recruits U2 small nuclear ribonucleoprotein to the branch site in the first ATP-dependent step in spliceosome assembly.

GENETICS
  #map_position 2 left arm
  SUMMARY #length 264 #molecular-weight 29855 #checksum 4187

Query Match 1.6%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 193 RARSRSRS 200
  |||||
  QY 148 RARSRSRS 155

RESULT 6
ENTRY A70039 #type complete
TITLE ABC transporter (ATP-binding protein) homolog yvfr - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS A70039
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bortoso, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borst, R.; Bouslier, L.; Brans, A.; Braun, M.; Briganti, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Choi, S.K.; Codani, J.U.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Diesterheft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Faurer, D.; Fitts, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Gilm, S.Y.; Glaeser, P.; Goffeau, A.; Golligorsky, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koertter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, V.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,

```

ENTRY TITLE ORGANISM DATE	RESULT 2	148723 #type complete ciliary neurotrophic factor - mouse #formal_name Mus musculus #common_name house mouse 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Feb-1997
Db	241	PRGCLIIINNHPEOMPTRNSTKDXKNLNLNPEOMGYATYCKDNLGRMLLTIRPFK 3000
Qy	241	PRGCLIIINNHPEOMPTRNSTKADKNLNLNPEOMGYATYCKDNLGRMLLTIRPFK 3000
Db	301	HESHGDSAILVILSHGEEENVIIAGVDIPISITHETIYDLLNANAPRLANKPIYFVQACRG 3600
Qy	301	HESHGDSAILVILSHGEEENVIIAGVDIPISITHETIYDLLNANAPRLANKPIYFVQACRG 3600
Db	361	ERRDNGEPVLDSVDGVAFLRKGMDNRDGLFNFLLGCVRPVOQOQWKKRKSQADILATYA 4200
Qy	361	ERRDNGEPVLDSVDGVAFLRKGMDNRDGLFNFLLGCVRPVOQOQWKKRKSQADILATYA 4200
Db	421	TTAQYVSMRNSARQSWFLQACEVFSTHARDMDVYELLTEVNNKRVACGPTQSGSNITLKO 4800
Qy	421	TTAQYVSMRNSARQSWFLQACEVFSTHARDMDVYELLTEVNNKRVACGPTQSGSNITLKO 4800
Db	481	MPEMTSRLLKKFFWMPPEARNSAV 503
Qy	481	MPEMTSRLLKKFFWMPPEARNSAV 503

```

#authors          Sactome, Y.; Walter, C.G.; Hirst, D
#journal          Gene (1995) 152:233-238
#title            A widely expressed novel C2H2 zinc-finger protein with
                  multiple consensus phosphorylation sites is conserved in
                  mouse and man.
#cross-references MUID:95137394
#accession        148723
#status           preliminary; translated from GB/EMBL/DBJ
#molecule_type  mRNA
#residues         1-198 ##label RES
#cross-references EMBL:U05342; NID:g453372; PID:g453374
CLASSIFICATION   #superfamily ciliary neurotrophic factor
SUMMARY          #length 198 #molecular-weight 22587 #checksum 8760

Query Match      1.68; Score 8; DB 2; Length 198;
Best Local Similarity 100.08; Pred. No. 5,18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      50 LDSYDGVF 57
        |||||
Qy      370 LDSYDGVF 377

RESULT      3
ENTRY       UNRBCF #type complete
TITLE       ciliary neurotrophic factor - rabbit
ALTERNATE_NAMES CNMF: Survival factor
ORGANISM    #formal_name Oryctolagus cuniculus #common_name domestic
              rabbit
DATE        31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
              30-Jun-1993
ACCESSIONS  A40082; B40082; C40082; A36286
REFERENCE   A40082
#authors    Lih, L.F.H.; Mismar, D.; Lih, J.D.; Armes, L.G.; Butler III,
              E.T.; Vannice, J.L.; Collins, F.
              Science (1989) 246:1023-1025
#journal    Purification, cloning, and expression of ciliary neurotrophic
#title      factor (CNMF).
#cross-references MUID:90069557
#accession  A40082
#molecule_type mRNA
#residues  1-199 ##label LII
#cross-references GB:M29828

```

827	6	1.2	1206	2	S72620	probable reverse tran	2.14e+02	900	1.2	3084	1	MMMSA	laminin alpha-1 chain	2.14e+02
828	6	1.2	1209	1	DNBEC4	DNA-binding protein -	2.14e+02	901	1.2	3144	2	A46068	Huntington disease-as	2.14e+02
829	6	1.2	1219	2	I61713	co-repressor protein	2.14e+02	902	1.2	3175	1	RNMVEV	genome polyprotein -	2.14e+02
830	6	1.2	1222	2	S40977	hypothetical protein	2.14e+02	903	1.2	3186	6	GNMVPD	genome polyprotein -	2.14e+02
831	6	1.2	1224	2	S73171	DNA-directed RNA poly	2.14e+02	904	1.2	3219	1	S43048	polyketide synthase t	2.14e+02
832	6	1.2	1228	2	I40468	surface layer protein	2.14e+02	905	1.2	3924	2	S37431	ankyrin 2, neuronal 1	2.14e+02
833	6	1.2	1229	2	A56068	co-repressor protein	2.14e+02	906	1.2	3972	2	S75251	hypothetical protein	2.14e+02
834	6	1.2	1232	2	D64413	cobalamin biosynthesi	2.14e+02	907	1.2	4391	2	A38096	perlecan precursor -	2.14e+02
835	6	1.2	1254	1	A32586	DNA-directed DNA poly	2.14e+02	908	1.2	4427	2	PN0637	polyketide synthase (2.14e+02
836	6	1.2	1259	2	S25954	gene alpha intron 2 pr	2.14e+02	909	1.2	4563	1	LPN05	apollipoprotein B-100	2.14e+02
837	6	1.2	1272	2	S60999	ubiquitin-specific pr	2.14e+02	910	1.2	4861	2	S71752	glant lipoprotein p619 -	2.14e+02
838	6	1.2	1287	1	S55954	viral mRNA translatio	2.14e+02	911	1.0	13	2	B52245	h1stone H1.c - mouse	4.86e+03
839	6	1.2	1289	1	GUPPT4	proximal tail fiber p	2.14e+02	912	1.0	86	2	S35769	T-cell receptor alpha	4.86e+03
840	6	1.2	1290	2	S73653	DNA-directed RNA poly	2.14e+02	913	1.0	92	2	S65009	lg heavy chain variab	4.86e+03
841	6	1.2	1291	2	S02021	microfil polyprotein	2.14e+02	914	1.0	108	2	C26405	lg kappa chain V regi	4.86e+03
842	6	1.2	1292	2	F64237	DNA-directed RNA poly	2.14e+02	915	1.0	109	2	S77210	perlaibulin - Japanes	4.86e+03
843	6	1.2	1299	1	WMBEHS	membrane antigen p140	2.14e+02	916	1.0	110	1	JN0761	red pigment-concentra	4.86e+03
844	6	1.2	1316	2	S31146	DNA-directed RNA poly	2.14e+02	917	1.0	110	2	S57465	lg lambda chain V-J r	4.86e+03
845	6	1.2	1316	2	G70535	probable rproc protein	2.14e+02	918	1.0	114	1	B31848	calgranulin B - human	4.86e+03
846	6	1.2	1317	2	S75517	DNA-directed RNA poly	2.14e+02	919	1.0	120	2	A29350	retrovirus-related po	4.86e+03
847	6	1.2	1323	2	T00037	hypothetical protein	2.14e+02	920	1.0	120	2	B24775	T-cell receptor gamma	4.86e+03
848	6	1.2	1332	2	F69732	PBSX prophage ORF xkd	2.14e+02	921	1.0	146	2	S40338	lg kappa chain - huma	4.86e+03
849	6	1.2	1338	2	S64356	phosphoribosylformylg	2.14e+02	922	1.0	146	2	S31589	lg heavy chain V regi	4.86e+03
850	6	1.2	1364	2	T00250	MEGF2 protein - human	2.14e+02	923	1.0	146	2	S30434	lg heavy chain V-D-J	4.86e+03
851	6	1.2	1371	2	S77521	sensory transduction	2.14e+02	924	1.0	153	2	S68816	legumin beta chain -	4.86e+03
852	6	1.2	1381	2	S55619	capsid protein 25 - e	2.14e+02	925	1.0	159	1	S31700	lg heavy chain V regi	4.86e+03
853	6	1.2	1386	2	S73401	MG064 homolog R02-ort	2.14e+02	926	1.0	175	1	CHN051	hemoglobin beta chain	4.86e+03
854	6	1.2	1396	2	S36851	L-shaped tail fiber p	2.14e+02	927	1.0	199	2	S55660	gamma-crystallin 1-2	4.86e+03
855	6	1.2	1402	2	S42748	finger protein - frui	2.14e+02	928	1.0	179	2	I39526	capsid protein 65 - e	4.86e+03
856	6	1.2	1403	2	A47328	natural killer cell t	2.14e+02	929	1.0	249	2	A41858	hypothetical protein	4.86e+03
857	6	1.2	1405	1	S13421	polymeric globulin alph	2.14e+02	930	1.0	200	2	G64106	3-isopropylmalate deh	4.86e+03
858	6	1.2	1416	2	D71350	probable DNA-directed	2.14e+02	931	1.0	215	2	G49343	probable dibenzothio	4.86e+03
859	6	1.2	1512	2	S72354	genome polyprotein -	2.14e+02	932	1.0	217	2	I48902	homeobox protein Pmx	4.86e+03
860	6	1.2	1518	2	S37928	probable purine nucle	2.14e+02	933	1.0	228	2	D30857	hypothetical protein	4.86e+03
861	6	1.2	1520	2	T00273	hypothetical protein	2.14e+02	934	1.0	237	2	H69820	conserved hypothetical	4.86e+03
862	6	1.2	1536	2	S59841	4-alpha-glucanotransf	2.14e+02	935	1.0	239	2	C64188	arginine binding prot	4.86e+03
863	6	1.2	1571	2	T00062	hypothetical protein	2.14e+02	936	1.0	245	2	A41858	o1 - Pseudomonas sp	4.86e+03
864	6	1.2	1574	2	G70466	DNA-directed RNA poly	2.14e+02	937	1.0	252	2	F65084	hypothetical protein -	4.86e+03
865	6	1.2	1592	2	S48933	probable transport pr	2.14e+02	938	1.0	265	2	S53614	insulin-like growth f	4.86e+03
866	6	1.2	1603	2	S17983	gene posterior sex co	2.14e+02	939	1.0	266	2	A35037	p15 system, n-acetyl	4.86e+03
867	6	1.2	1676	2	E71410	probable centromere p	2.14e+02	940	1.0	267	2	G65103	N-hydroxyarylamino O-	4.86e+03
868	6	1.2	1676	2	A56508	anucleate primary ste	2.14e+02	941	1.0	281	2	A38090	Na+/K+-exchanging ATP	4.86e+03
869	6	1.2	1683	2	S56811	probable membrane pro	2.14e+02	942	1.0	305	1	PMRYNB	heterogeneous ribonuc	4.86e+03
870	6	1.2	1722	2	I78879	retinoblastoma bindin	2.14e+02	943	1.0	365	2	A34840	surfactant protein D	4.86e+03
871	6	1.2	1729	2	S57596	ribosomal RNA process	2.14e+02	944	1.0	365	2	S33603	conglutinin precursor	4.86e+03
872	6	1.2	1736	2	A47747	light junction protei	2.14e+02	945	1.0	371	2	JN0450	alattine dehydrogenase	4.86e+03
873	6	1.2	1742	2	S24600	F44E2.1 protein - Cae	2.14e+02	946	1.0	372	2	B34261	ubiquinol--cytochrome	4.86e+03
874	6	1.2	1745	2	S44816	hypothetical protein	2.14e+02	947	1.0	379	2	I48133	rhizopuspepsin (EC 3.	4.86e+03
875	6	1.2	1770	2	A71517	laminin beta-1 chain	2.14e+02	948	1.0	393	1	A26681	probable phosphoester	4.86e+03
876	6	1.2	1805	2	MMHUB1	hypothetical protein	2.14e+02	949	1.0	408	2	A69819	pol polyprotein - min	4.86e+03
877	6	1.2	1805	2	A64224	hypothetical protein	2.14e+02	950	1.0	409	2	S12588	gamma-Dm protein - Rhizob	4.86e+03
878	6	1.2	1818	2	S73852	DNA-directed DNA poly	2.14e+02	951	1.0	412	2	A30605	tyrosine--tRNA ligase	4.86e+03
879	6	1.2	1894	2	T02155	hypothetical protein	2.14e+02	952	1.0	419	1	SYBSYF	dihydroorotase - Aqu	4.86e+03
880	6	1.2	1922	2	T00637	cag pathogenicity isl	2.14e+02	953	1.0	422	2	C70370	tailless (Tli) protei	4.86e+03
881	6	1.2	1927	2	G64585	probable polyketide s	2.14e+02	954	1.0	449	2	C39926	IL protein - human he	4.86e+03
882	6	1.2	1937	2	T03324	NUMA protein - human	2.14e+02	955	1.0	450	2	B47265	perlecan intermedia	4.86e+03
883	6	1.2	2101	2	A42184	NUMA protein - human	2.14e+02	956	1.0	458	2	B33560	Cab-binding protein a	4.86e+03
884	6	1.2	2115	2	S23647	SEC16 protein - yeast	2.14e+02	957	1.0	459	2	H64024	helicase (EC 3.6.1.-)	4.86e+03
885	6	1.2	2195	2	S61103	genome polyprotein 1	2.14e+02	958	1.0	468	2	JN0016	probable gnd protein	4.86e+03
886	6	1.2	2252	2	S06188	acetyl-CoA carboxylas	2.14e+02	959	1.0	469	1	NMMSA	hypothetical protein	4.86e+03
887	6	1.2	2257	2	A57710	genome polyprotein 1	2.14e+02	960	1.0	475	1	Z4BP74	photosystem II chloro	4.86e+03
888	6	1.2	2264	1	GNVTB	genome polyprotein 1	2.14e+02	961	1.0	485	2	D70664	cytochrome P450 1A, h	4.86e+03
889	6	1.2	2325	2	A61208	chondroitin sulfate p	2.14e+02	962	1.0	491	2	S73110	cytochrome d ubiquino	4.86e+03
890	6	1.2	2489	2	S59782	probable membrane pro	2.14e+02	963	1.0	504	2	B40829	keratin, 54K type I c	4.86e+03
891	6	1.2	2515	2	S47008	transcription factor	2.14e+02	964	1.0	508	1	QUTV6A	Hr-transporting Arp s	4.86e+03
892	6	1.2	2529	2	A56923	transcription factor	2.14e+02	965	1.0	514	2	A36793	cyclin-dependent kina	4.86e+03
893	6	1.2	2549	2	A54837	rapamycin/FKBP2 targ	2.14e+02	966	1.0	521	2	S53138		
894	6	1.2	2549	2	S45340	FKBP-rapamycin-associ	2.14e+02	967	1.0	523	2	D64809		
895	6	1.2	2594	2	A35774	kinase-related protei	2.14e+02	968	1.0	526	1	KRB0V1		
896	6	1.2	2763	1	WZBE22	gene 22 protein - hum	2.14e+02	969	1.0	536	2	A40829		
897	6	1.2	2893	1	A64556	toxin-like outer memb	2.14e+02	970	1.0	545	1	PMBYA		
898	6	1.2	3063	2	S55505	fatty-acid synthase (2.14e+02	971	1.0	555	2	S50157		
899	6	1.2	3075	2	S14458	laminin alpha-1 chain	2.14e+02	972	1.0					

681	6	1.2	638	2	B35816	transcription regulat	2.14e+02	754	6	1.2	814	2	I39627	nicotine dehydrogenas	2.14e+02
682	6	1.2	639	2	H70661	probable diag protein	2.14e+02	755	6	1.2	815	2	B56708	extracellular signal-	2.14e+02
683	6	1.2	639	2	G69157	exonuclease ABC chain	2.14e+02	756	6	1.2	816	2	A71006	hypothetical protein	2.14e+02
684	6	1.2	646	2	A37086	beta-galactosidase (E	2.14e+02	757	6	1.2	822	2	J70968	1,4-alpha-glucan bran	2.14e+02
685	6	1.2	647	2	A35816	transcription regulat	2.14e+02	758	6	1.2	825	2	B64689	adenine specific DNA	2.14e+02
686	6	1.2	652	2	B29900	fasciclin I precursor	2.14e+02	759	6	1.2	840	2	A27832	cell division control	2.14e+02
687	6	1.2	657	2	E70529	hypothetical protein	2.14e+02	760	6	1.2	858	2	SC30571	DNA topoisomerase (Ar	2.14e+02
688	6	1.2	661	2	B71447	probable GLABRA2 - Ar	2.14e+02	761	6	1.2	859	1	VCLJEL1	env polyprotein precu	2.14e+02
689	6	1.2	662	2	E71731	exonuclease ABC chain	2.14e+02	762	6	1.2	859	1	VCLJ22	env polyprotein precu	2.14e+02
690	6	1.2	662	2	E64650	acetyl-coA synthetase	2.14e+02	763	6	1.2	859	1	VCLJEV	env polyprotein precu	2.14e+02
691	6	1.2	663	2	S20877	dnak-type molecular c	2.14e+02	764	6	1.2	859	1	VCLJEW	env polyprotein precu	2.14e+02
692	6	1.2	668	2	H71312	probable ATP-dependen	2.14e+02	765	6	1.2	859	1	VCLJEE	env polyprotein precu	2.14e+02
693	6	1.2	668	2	SC9836	hypothetical protein	2.14e+02	766	6	1.2	859	1	VCLJEE	env polyprotein precu	2.14e+02
694	6	1.2	669	2	S64795	suppressor protein FP	2.14e+02	767	6	1.2	860	1	VCLJEA	env polyprotein precu	2.14e+02
695	6	1.2	673	2	S59263	probable membrane pro	2.14e+02	768	6	1.2	860	1	VCLJEA	env polyprotein precu	2.14e+02
696	6	1.2	673	1	TVBER5	74 kD protein kinase	2.14e+02	769	6	1.2	862	1	E35905	ATP-dependent Clp pro	2.14e+02
697	6	1.2	674	2	T01309	probable serine/threo	2.14e+02	770	6	1.2	864	2	JH0438	penicillin-binding pr	2.14e+02
698	6	1.2	676	1	WMBEX6	U6 protein - human h	2.14e+02	771	6	1.2	867	2	T00118	hrsh2 protein - sea s	2.14e+02
699	6	1.2	677	2	S33664	flagella-associated p	2.14e+02	772	6	1.2	868	2	S65186	NiP80 protein - yeast	2.14e+02
700	6	1.2	680	4	I38491	nucleophosmin/naplas	2.14e+02	773	6	1.2	878	2	S20486	paramyosin - fruit fl	2.14e+02
701	6	1.2	681	2	A36500	transferrin precursor	2.14e+02	774	6	1.2	879	2	S22028	telomerase catalytic	2.14e+02
702	6	1.2	682	2	S40037	pili protein - Pseudo	2.14e+02	775	6	1.2	884	2	S53396	telomerase catalytic	2.14e+02
703	6	1.2	683	2	D71680	ribonuclease E (rne)	2.14e+02	776	6	1.2	885	1	WMBX3L	ribonucleoside-diphos	2.14e+02
704	6	1.2	684	2	E64496	eIF-4A family probabl	2.14e+02	777	6	1.2	892	2	S20602	probable transposase	2.14e+02
705	6	1.2	690	2	S75067	H+/K+-exchanging ATPa	2.14e+02	778	6	1.2	902	2	S26002	gene coxi intron 1 pr	2.14e+02
706	6	1.2	691	1	A36295	heat shock transcript	2.14e+02	779	6	1.2	905	2	H71731	DNA gyrase chain A (E	2.14e+02
707	6	1.2	695	2	I54325	gene XE7 protein - hu	2.14e+02	780	6	1.2	910	2	A65000	NADH dehydrogenase A	2.14e+02
708	6	1.2	697	2	S72353	capsid protein - Sacc	2.14e+02	781	6	1.2	917	1	VGBEBH	glycoprotein B precu	2.14e+02
709	6	1.2	700	2	A61527	stomoxystoxin beta cha	2.14e+02	782	6	1.2	921	2	A33718	retinoblastoma protel	2.14e+02
710	6	1.2	701	2	S61239	hypothetical protein	2.14e+02	783	6	1.2	922	2	S54342	retinoblastoma protel	2.14e+02
711	6	1.2	701	2	F70155	Na+/H+ antiporter (na	2.14e+02	784	6	1.2	925	1	RBHU	retinoblastoma-associ	2.14e+02
712	6	1.2	702	1	A48562	coat protein - San Mi	2.14e+02	785	6	1.2	938	2	S55080	hypothetical protein	2.14e+02
713	6	1.2	702	2	A36319	carcinoembryonic anti	2.14e+02	786	6	1.2	940	2	A40985	Bicaudal-C - fruit fl	2.14e+02
714	6	1.2	705	2	S55430	conserved hypotheticala	2.14e+02	787	6	1.2	948	2	T03306	protection - fruit fly	2.14e+02
715	6	1.2	706	2	T01351	subtilisin-like prote	2.14e+02	788	6	1.2	948	2	T03306	PSD-95/SAR90-associ	2.14e+02
716	6	1.2	706	2	S61717	probable membrane pro	2.14e+02	789	6	1.2	948	2	A57640	retinoblastoma protel	2.14e+02
717	6	1.2	707	2	A42322	ornithine decarboxyla	2.14e+02	790	6	1.2	959	2	B71405	probable kinesin - Ar	2.14e+02
718	6	1.2	723	2	S51788	malate synthase (EC 4	2.14e+02	791	6	1.2	962	2	S03818	carboxymethylcellulas	2.14e+02
719	6	1.2	727	2	S53707	translation initiatio	2.14e+02	792	6	1.2	964	2	S06028	gene suppressor-of-wh	2.14e+02
720	6	1.2	727	2	H69724	DNA topoisomerase III	2.14e+02	793	6	1.2	979	2	A70848	probable membrane pro	2.14e+02
721	6	1.2	729	2	S76065	hypothetical protein	2.14e+02	794	6	1.2	979	2	JC2349	protein-tyrosine phos	2.14e+02
722	6	1.2	731	2	D71332	probable DNA topoisom	2.14e+02	795	6	1.2	996	2	I48721	PTP 35 protein - mous	2.14e+02
723	6	1.2	734	2	C69691	GTP pyrophosphokinase	2.14e+02	796	6	1.2	1005	2	A64465	hypothetical protein	2.14e+02
724	6	1.2	739	2	I40715	malate synthase (EC 4	2.14e+02	797	6	1.2	1009	2	S64734	retrovirus-related po	2.14e+02
725	6	1.2	739	1	J01893	80.7K alpha trans-ind	2.14e+02	798	6	1.2	1020	1	QFHHH	neurofilament triplet	2.14e+02
726	6	1.2	741	2	E70722	probable glcB protein	2.14e+02	799	6	1.2	1026	2	S51432	hypothetical protein	2.14e+02
727	6	1.2	742	2	S12533	zeta protein - mouse	2.14e+02	800	6	1.2	1028	2	T03516	probable outer membra	2.14e+02
728	6	1.2	743	1	TNBEI1	80.7K alpha trans-ind	2.14e+02	801	6	1.2	1030	2	S43311	hypothetical ATP-binding	2.14e+02
729	6	1.2	743	1	DECHE	glutamate dehydrogena	2.14e+02	802	6	1.2	1046	2	S56026	hypothetical protein	2.14e+02
730	6	1.2	743	2	S38143	hypothetical protein	2.14e+02	803	6	1.2	1055	2	JC5216	type I site-specific	2.14e+02
731	6	1.2	746	2	S67203	probable membrane pro	2.14e+02	804	6	1.2	1059	1	KUHH	ferrooxidase (BC 1.16,	2.14e+02
732	6	1.2	747	2	I48294	gene Clc4 protein -	2.14e+02	805	6	1.2	1072	2	A37221	neurofilament triplet	2.14e+02
733	6	1.2	747	1	A57107	kinesin-related prote	2.14e+02	806	6	1.2	1078	2	S77162	DNA topoisomerase (Ar	2.14e+02
734	6	1.2	748	2	E71927	csg island protein, D	2.14e+02	807	6	1.2	1081	1	WZBPA6	119K DNA helicase/prt	2.14e+02
735	6	1.2	749	2	S50095	splicing regulator Ce	2.14e+02	808	6	1.2	1086	1	G02257	NAD(P)+ transhydrogen	2.14e+02
736	6	1.2	752	2	E40780	translatation elongator	2.14e+02	809	6	1.2	1086	1	DEBOXM	NAD(P)+ transhydrogen	2.14e+02
737	6	1.2	754	2	JC5314	CBC28/cdc2-like kinas	2.14e+02	810	6	1.2	1086	1	S54876	NAD(P)+ transhydrogen	2.14e+02
738	6	1.2	758	3	T02925	protoporphyrin IX mag	2.14e+02	811	6	1.2	1087	1	QFMSH	neurofilament triplet	2.14e+02
739	6	1.2	760	2	S33752	dipeptidyl-peptidase	2.14e+02	812	6	1.2	1093	2	I38533	AF17 protein - human	2.14e+02
740	6	1.2	773	2	A36932	iron(II) transposr (E	2.14e+02	813	6	1.2	1095	2	T00329	hypothetical protein	2.14e+02
741	6	1.2	774	2	A28392	penicillin amidase (S	2.14e+02	814	6	1.2	1129	2	A29631	phytochrome - oat	2.14e+02
742	6	1.2	776	2	S03611	outer layer protein V	2.14e+02	815	6	1.2	1129	2	S00097	phytochrome 3 - oat	2.14e+02
743	6	1.2	776	1	VXP851	outer layer protein V	2.14e+02	816	6	1.2	1139	2	S00096	phytochrome 3 - oat	2.14e+02
744	6	1.2	778	2	E71263	probable tek protein	2.14e+02	817	6	1.2	1130	2	A48843	MHC class IIT transact	2.14e+02
745	6	1.2	779	2	ES1413	probable kinase Draf-1	2.14e+02	818	6	1.2	1132	2	A35098	MHC class IIT histoco	2.14e+02
746	6	1.2	781	1	TVFDFP	protein kinase Draf-1	2.14e+02	819	6	1.2	1132	2	S60433	probable membrane pro	2.14e+02
747	6	1.2	785	2	D71653	cell surface antigen	2.14e+02	820	6	1.2	1139	2	A54962	sterol regulatory ele	2.14e+02
748	6	1.2	790	2	S67803	probable membrane pro	2.14e+02	821	6	1.2	1141	2	A54962	sterol regulatory ele	2.14e+02
749	6	1.2	801	2	PC6010	RNA helicase Gu - hum	2.14e+02	822	6	1.2	1142	2	S50632	protein kinase Pak1 (2.14e+02
750	6	1.2	802	2	S43882	PMH protein, splice f	2.14e+02	823	6	1.2	1148	2	T00016	minor outer capsid pr	2.14e+02
751	6	1.2	802	2	S42518	PMH protein, splice f	2.14e+02	824	6	1.2	1164	2	G71827	hypothetical protein	2.14e+02
752	6	1.2	808	2	JQ2205	U547H protein - Marek	2.14e+02	825	6	1.2	1195	2	C64440	DNA repair protein RA	2.14e+02
753	6	1.2	812	1	PLBO	plasmin (EC 3.4.21.7)	2.14e+02	826	6	1.2	1199	2	G69698	DNA polymerase (beta)	2.14e+02

535	6	1.2	463	1	WZBYS	arginosuccinate lya	2.14e+02	608	6	1.2	531	2	S76041	hypothetical protein	2.14e+02
536	6	1.2	464	2	S16954	retinoblastoma-binding	2.14e+02	609	6	1.2	532	2	S33758	flavin-containing mon	2.14e+02
537	6	1.2	465	2	UC1318	triacylglycerol lipas	2.14e+02	610	6	1.2	532	1	A33768	dimethylamline mono	2.14e+02
538	6	1.2	466	2	JN0765	alpha-1C-adrenergic r	2.14e+02	611	6	1.2	534	2	S62572	hypothetical protein	2.14e+02
539	6	1.2	466	2	I52862	alpha-1C-adrenergic re	2.14e+02	612	6	1.2	534	2	S55635	DNA helicase-primase	2.14e+02
540	6	1.2	466	2	I57959	alpha-1C-adrenergic re	2.14e+02	613	6	1.2	535	1	A35182	dimethylamline mono	2.14e+02
541	6	1.2	466	2	A35375	alpha-1-adrenergic re	2.14e+02	614	6	1.2	536	1	A35427	dimethylamline mono	2.14e+02
542	6	1.2	467	1	PMB5BF	H+-transporting ATP s	2.14e+02	615	6	1.2	539	2	T02705	hypothetical protein	2.14e+02
543	6	1.2	467	1	VHNZ	nucleocapsid protein	2.14e+02	616	6	1.2	542	2	S45557	hypothetical resb hom	2.14e+02
544	6	1.2	467	2	S30839	UTR2 protein - yeast	2.14e+02	617	6	1.2	543	1	KIHUPL	pyruvate kinase (EC 2	2.14e+02
545	6	1.2	467	2	S38458	replication protein A	2.14e+02	618	6	1.2	543	2	A23612	pyruvate kinase (EC 2	2.14e+02
546	6	1.2	468	2	S20951	Na+/H+-exchanging pro	2.14e+02	619	6	1.2	543	1	KIRPPL	pyruvate kinase (EC 2	2.14e+02
547	6	1.2	469	2	JC5741	membrane-bound protein	2.14e+02	620	6	1.2	546	2	S67292	probable membrane pro	2.14e+02
548	6	1.2	469	2	S53810	paracrystalline surfa	2.14e+02	621	6	1.2	547	1	ERADF4	60.5K fiber protein -	2.14e+02
549	6	1.2	469	2	S17726	H+-transporting ATP s	2.14e+02	622	6	1.2	547	2	I37577	islet cell antigen 51	2.14e+02
550	6	1.2	470	2	A65168	hypothetical 49.9 KD	2.14e+02	623	6	1.2	549	2	S47739	probable alpha, alpha-	2.14e+02
551	6	1.2	470	2	H70795	hypothetical protein	2.14e+02	624	6	1.2	553	2	D69162	pyruvate dehydrogenas	2.14e+02
552	6	1.2	470	2	C70672	hypothetical protein	2.14e+02	625	6	1.2	554	2	JU0188	membrane protein 4.1	2.14e+02
553	6	1.2	471	2	S69435	hypothetical protein	2.14e+02	626	6	1.2	554	2	C64762	probable monooxygenas	2.14e+02
554	6	1.2	472	1	PMB5BM	H+-transporting ATP s	2.14e+02	627	6	1.2	557	2	S62522	hypothetical protein	2.14e+02
555	6	1.2	474	2	S00880	nitrogenase molybdenu	2.14e+02	628	6	1.2	558	1	DEHE	glutamate dehydrogena	2.14e+02
556	6	1.2	475	2	S20250	splicing factor U2AF	2.14e+02	629	6	1.2	558	1	S03707	glutamate dehydrogena	2.14e+02
557	6	1.2	476	2	JC4646	bone morphogenetic pr	2.14e+02	630	6	1.2	558	1	S16239	glutamate dehydrogena	2.14e+02
558	6	1.2	476	2	C64601	bone morphogenetic pr	2.14e+02	631	6	1.2	562	1	ERADN2	60.5K fiber protein -	2.14e+02
559	6	1.2	477	2	S71323	alpha-1A adrenergic r	2.14e+02	632	6	1.2	565	2	T01116	carrot B2 protein hom	2.14e+02
560	6	1.2	478	2	JC4838	bone morphogenetic pr	2.14e+02	633	6	1.2	566	1	HMTVNI	hemagglutinin precurs	2.14e+02
561	6	1.2	478	2	A32555	major protozoite surfa	2.14e+02	634	6	1.2	566	1	HMTVD1	hemagglutinin precurs	2.14e+02
562	6	1.2	479	1	VGBEF2	glycoprotein F - huma	2.14e+02	635	6	1.2	566	1	HMTIVR	hemagglutinin precurs	2.14e+02
563	6	1.2	481	2	S69808	lincomycin resistance	2.14e+02	636	6	1.2	566	1	HMTIVR	hemagglutinin precurs	2.14e+02
564	6	1.2	482	3	A43828	serine carboxypeptida	2.14e+02	637	6	1.2	566	1	B41648	hemagglutinin precurs	2.14e+02
565	6	1.2	485	2	JU0300	X-His dipeptidase (EC	2.14e+02	638	6	1.2	566	2	S69888	hemagglutinin precurs	2.14e+02
566	6	1.2	490	2	S70352	protein kinase CLK3 (2.14e+02	639	6	1.2	566	2	S69889	hemagglutinin precurs	2.14e+02
567	6	1.2	490	2	S53639	protein kinase CLK3 (2.14e+02	640	6	1.2	566	1	HMTV	hemagglutinin precurs	2.14e+02
568	6	1.2	490	2	E70649	probable regulatory p	2.14e+02	641	6	1.2	568	1	JC4570	gamma-glutamyltransfe	2.14e+02
569	6	1.2	490	2	S30917	bcl2 protein - Rhodop	2.14e+02	642	6	1.2	568	1	A05225	nodu protein - Bradyr	2.14e+02
570	6	1.2	492	2	S22646	splicing factor U2AF	2.14e+02	643	6	1.2	569	2	S27495	L-Ornithine N5-oxygen	2.14e+02
571	6	1.2	492	2	S32951	regulatory protein hu	2.14e+02	644	6	1.2	570	2	A47266	axl 1-like protein -	2.14e+02
572	6	1.2	495	2	S00098	phytochrome 5 - oat (2.14e+02	645	6	1.2	573	2	T02405	probable pyrophosphat	2.14e+02
573	6	1.2	496	2	B64638	conserved hypothetical	2.14e+02	646	6	1.2	574	1	KIHUPR	pyruvate kinase (EC 2	2.14e+02
574	6	1.2	496	2	A49418	spermatogenesis facto	2.14e+02	647	6	1.2	574	1	KIHUPR	pyruvate kinase (EC 2	2.14e+02
575	6	1.2	499	2	A38891	cytochrome-c oxidase	2.14e+02	648	6	1.2	578	2	JC1385	tyrosine kinase (EC 2	2.14e+02
576	6	1.2	499	2	S65657	alpha-1C-adrenergic r	2.14e+02	649	6	1.2	579	2	A55369	tyrosine kinase (EC 2	2.14e+02
577	6	1.2	500	2	A53658	prostaglandin-I synth	2.14e+02	650	6	1.2	581	2	E69322	dolichol-P-glucose sy	2.14e+02
578	6	1.2	500	2	JC2231	prostaglandin-I synth	2.14e+02	651	6	1.2	583	2	S50959	probable membrane pro	2.14e+02
579	6	1.2	501	2	S44258	sucrose-6-phosphate h	2.14e+02	652	6	1.2	583	2	S19476	hypothetical protein	2.14e+02
580	6	1.2	502	2	F64543	conserved hypothetical	2.14e+02	653	6	1.2	584	2	G70804	hypothetical glycine-	2.14e+02
581	6	1.2	502	2	E71963	probable cardiolipin	2.14e+02	654	6	1.2	588	2	A49618	probable ataxia-felan	2.14e+02
582	6	1.2	503	2	S23243	glutamate dehydrogena	2.14e+02	655	6	1.2	590	1	A38119	cell division protein	2.14e+02
583	6	1.2	504	1	DEBOE	hypothetical protein	2.14e+02	656	6	1.2	595	2	F70402	penicillin binding pr	2.14e+02
584	6	1.2	505	2	T00723	hypothetical protein	2.14e+02	657	6	1.2	599	2	H71932	hypothetical protein	2.14e+02
585	6	1.2	506	2	S67256	probable pyruvate kin	2.14e+02	658	6	1.2	599	2	H64579	GTP-binding protein	2.14e+02
586	6	1.2	508	1	PW2MAM	H+-transporting ATP s	2.14e+02	659	6	1.2	601	2	S12004	tyramine receptor - f	2.14e+02
587	6	1.2	509	2	S18872	legumin-like storage	2.14e+02	660	6	1.2	601	2	G71666	nicotinic regulation p	2.14e+02
588	6	1.2	510	2	S41307	transcription initiat	2.14e+02	661	6	1.2	601	2	JH0170	octopamine receptor t	2.14e+02
589	6	1.2	510	2	S56315	hypothetical ABC tran	2.14e+02	662	6	1.2	603	2	H70156	exonuclease ABC, sub	2.14e+02
590	6	1.2	511	2	C68199	phenylalanine-tRNA t	2.14e+02	663	6	1.2	603	2	A47545	protein kinase (EC 2	2.14e+02
591	6	1.2	513	2	A70378	histidine kinase sens	2.14e+02	664	6	1.2	603	2	A54596	protein kinase - mous	2.14e+02
592	6	1.2	513	2	D69016	serine-tRNA ligase (2.14e+02	665	6	1.2	603	2	S34130	serine/threonine-spec	2.14e+02
593	6	1.2	514	1	W2WL5	E2 protein - human pa	2.14e+02	666	6	1.2	603	2	A22282	DNA primase (EC 2.7.7	2.14e+02
594	6	1.2	515	2	JC1525	alpha-1B-adrenergic r	2.14e+02	667	6	1.2	603	2	S76959	GTP-binding membrane	2.14e+02
595	6	1.2	515	2	A40491	alpha-1B-adrenergic re	2.14e+02	668	6	1.2	607	2	S42639	ATP-dependent RNA hel	2.14e+02
596	6	1.2	515	1	VCLJB	env polypeptide - bov	2.14e+02	669	6	1.2	609	2	A64432	nodulation factor pro	2.14e+02
597	6	1.2	515	2	S59811	vacuolar segregation	2.14e+02	670	6	1.2	613	2	T00077	gag-1-like protein - Ch	2.14e+02
598	6	1.2	516	2	JC3332	alpha-1B-adrenergic r	2.14e+02	671	6	1.2	614	2	A25707	U1 snRNP 70K protein	2.14e+02
599	6	1.2	517	2	A45121	hypothetical protein r	2.14e+02	672	6	1.2	616	2	C69226	type I restriction mo	2.14e+02
600	6	1.2	521	2	F70772	hypothetical protein	2.14e+02	673	6	1.2	618	2	A36907	p90c1 SI SIA subunit	2.14e+02
601	6	1.2	523	2	S47758	hypothetical 59.4K pr	2.14e+02	674	6	1.2	619	2	A36682	72K mitochondrial out	2.14e+02
602	6	1.2	525	2	JN0443	transcription initiat	2.14e+02	675	6	1.2	622	2	S62532	hypothetical protein	2.14e+02
603	6	1.2	527	1	KIBYD	aspartate kinase (EC	2.14e+02	676	6	1.2	627	2	T02846	hypothetical protein	2.14e+02
604	6	1.2	528	2	JN0445	transcription initiat	2.14e+02	677	6	1.2	630	2	B24584	transpositional regulat	2.14e+02
605	6	1.2	528	2	S19366	hypothetical protein	2.14e+02	678	6	1.2	632	2	A71259	probable dicarboxylat	2.14e+02
606	6	1.2	529	1	VGNVAC	major envelope glycop	2.14e+02	679	6	1.2	635	2	E71272	probable glucosamine-	2.14e+02
607	6	1.2	530	2	S12320	PRP9 protein - yeast	2.14e+02	680	6	1.2	637	2	JC5608	dnak-type molecular c	2.14e+02

389	6	1.2	346	2	E64002	hypothetical protein	2.14e+02	462	6	1.2	401	2	A47258	interleukin-1 beta-co	2.14e+02
390	6	1.2	347	1	C6137	opsin, violet-sensiti	2.14e+02	463	6	1.2	401	2	S63227	hypothetical protein	2.14e+02
391	6	1.2	347	2	S09274	Ig alpha chain C regi	2.14e+02	464	6	1.2	401	2	S77167	isoprenicillin N epim	2.14e+02
392	6	1.2	349	2	A34815	carcinoembryonic anti	2.14e+02	465	6	1.2	402	2	UC3382	ribosomal protein L3	2.14e+02
393	6	1.2	350	2	A69834	conserved hypothetical	2.14e+02	466	6	1.2	402	2	A46495	IL-1 beta convertase	2.14e+02
394	6	1.2	350	2	B85835	A/G-specific adenine	2.14e+02	467	6	1.2	403	1	A53396	flavohemoglobin - Alc	2.14e+02
395	6	1.2	352	2	150047	Rhodopsin - Mexican t	2.14e+02	468	6	1.2	404	2	C70610	probable glucose-1-ph	2.14e+02
396	6	1.2	352	2	S24559	Mnt-2 protein - fruit	2.14e+02	469	6	1.2	406	2	A70015	NADH dehydrogenase ho	2.14e+02
397	6	1.2	353	2	S16785	hemagglutinin - influ	2.14e+02	470	6	1.2	406	2	UC4600	isocitrate dehydrogen	2.14e+02
398	6	1.2	354	2	A35788	hemagglutinin - influ	2.14e+02	471	6	1.2	407	2	UC5632	hemocyanin dioxigen-b	2.14e+02
399	6	1.2	354	2	E69499	iron-sulfur binding r	2.14e+02	472	6	1.2	407	2	B70733	probable aminotransfe	2.14e+02
400	6	1.2	354	2	180165	class I histocompatib	2.14e+02	473	6	1.2	411	2	D69316	mRNA 3'-end processin	2.14e+02
401	6	1.2	354	2	A48049	coproporphyrinogen ox	2.14e+02	474	6	1.2	408	2	G70522	probable transposase	2.14e+02
402	6	1.2	354	2	152444	coproporphyrinogen ox	2.14e+02	475	6	1.2	409	2	I46880	transcription factor	2.14e+02
403	6	1.2	355	2	S76940	hypothetical protein	2.14e+02	476	6	1.2	409	2	S60975	hypothetical protein	2.14e+02
404	6	1.2	355	2	F69298	conserved hypothetical	2.14e+02	477	6	1.2	409	2	S60988	hypothetical protein	2.14e+02
405	6	1.2	355	2	151319	RH2 opsin - green ano	2.14e+02	478	6	1.2	410	2	150494	serine proteinase inh	2.14e+02
406	6	1.2	355	2	A42347	opsin, green-sensitiv	2.14e+02	479	6	1.2	411	2	S37644	protein kinase MSK-1	2.14e+02
407	6	1.2	355	2	A46191	iodopsin homolog to	2.14e+02	480	6	1.2	412	2	S31505	serine proteinase inh	2.14e+02
408	6	1.2	356	2	G69529	hypothetical protein	2.14e+02	481	6	1.2	413	2	F69007	translational elongatio	2.14e+02
409	6	1.2	356	2	A47128	carinomycin 4-O-meth	2.14e+02	482	6	1.2	414	2	S48738	potassium channel - m	2.14e+02
410	6	1.2	358	2	F70583	hypothetical protein	2.14e+02	483	6	1.2	414	2	A53950	transcription factor	2.14e+02
411	6	1.2	358	2	S51263	probable galactosyltr	2.14e+02	484	6	1.2	415	2	158144	coriicolipin-releasi	2.14e+02
412	6	1.2	358	2	S68886	hemagglutinin precurs	2.14e+02	485	6	1.2	415	2	S39535	coriicolipin-releasi	2.14e+02
413	6	1.2	360	2	S65210	hypothetical protein	2.14e+02	486	6	1.2	415	2	A56042	mitogen-activated pro	2.14e+02
414	6	1.2	360	2	S74751	CDP-glucose 4,6-dehyd	2.14e+02	487	6	1.2	416	2	A70327	conserved hypothetical	2.14e+02
415	6	1.2	361	2	B46137	opsin, blue-sensitiv	2.14e+02	488	6	1.2	416	2	G02635	ICF-LAP6 - human	2.14e+02
416	6	1.2	362	2	S69197	oleoyl-lacetyl-carrier	2.14e+02	489	6	1.2	417	2	F64660	colicin tolerance-lik	2.14e+02
417	6	1.2	364	1	OOHUG	opsin, green-sensitiv	2.14e+02	490	6	1.2	417	2	G64614	conserved hypothetical	2.14e+02
418	6	1.2	364	1	J70741	GTP-binding protein 1	2.14e+02	491	6	1.2	417	2	G71854	probable tonb-indepen	2.14e+02
419	6	1.2	364	1	OOHUR	opsin, red-sensitiv	2.14e+02	492	6	1.2	417	2	H71898	hypothetical protein	2.14e+02
420	6	1.2	365	2	S63542	hypothetical protein	2.14e+02	493	6	1.2	417	2	B55473	early switch protein	2.14e+02
421	6	1.2	366	1	D2EC3B	DNA-directed DNA poly	2.14e+02	494	6	1.2	419	2	S68803	probable transcriptio	2.14e+02
422	6	1.2	367	1	A46355	site-specific DNA-met	2.14e+02	495	6	1.2	419	2	G70852	probable sers protein	2.14e+02
423	6	1.2	371	2	D71201	hypothetical protein	2.14e+02	496	6	1.2	419	2	B49418	spermatogenesis facto	2.14e+02
424	6	1.2	372	2	150492	hypothetical protein	2.14e+02	497	6	1.2	420	2	151666	Mel-1c receptor subty	2.14e+02
425	6	1.2	372	2	A33375	alpha-1-antitrypsin p	2.14e+02	498	6	1.2	420	2	E70914	probable lipD protein	2.14e+02
426	6	1.2	375	2	138879	lymph node homing rec	2.14e+02	499	6	1.2	422	1	A56674	paired box transcript	2.14e+02
427	6	1.2	376	2	B69125	corticotropin releasi	2.14e+02	500	6	1.2	423	2	A69367	translational elongatio	2.14e+02
428	6	1.2	376	2	S16386	hydrogenase expressio	2.14e+02	501	6	1.2	423	2	F64436	hypothetical protein	2.14e+02
429	6	1.2	376	2	A57511	hydrogenase expressio	2.14e+02	502	6	1.2	424	2	I38979	inward rectifier pota	2.14e+02
430	6	1.2	377	2	A57511	hydrogenase expressio	2.14e+02	503	6	1.2	424	2	B69210	inward rectifier pota	2.14e+02
431	6	1.2	377	2	S74628	interleukin-1 beta co	2.14e+02	504	6	1.2	425	2	S52852	conserved hypotherica	2.14e+02
432	6	1.2	377	2	S12190	hypothetical protein	2.14e+02	505	6	1.2	425	2	C64567	inward rectifier pota	2.14e+02
433	6	1.2	378	2	B69344	mobl protein - Rhinob	2.14e+02	506	6	1.2	425	2	I48970	fucoyltransferase -	2.14e+02
434	6	1.2	379	2	A70646	hypothetical protein	2.14e+02	507	6	1.2	426	2	D64870	potassium channel-lik	2.14e+02
435	6	1.2	380	2	C67416	probable PPE protein	2.14e+02	508	6	1.2	426	2	S19338	hypothetical protein	2.14e+02
436	6	1.2	381	2	E36607	conserved hypothetical	2.14e+02	509	6	1.2	429	2	S65656	alpha-1C-adrenergic r	2.14e+02
437	6	1.2	382	2	S03762	nucleocapsid protein	2.14e+02	510	6	1.2	433	2	B69495	acetylase (acn) homol	2.14e+02
438	6	1.2	382	2	S36476	nucleocapsid protein	2.14e+02	511	6	1.2	435	2	B69115	coenzyme F390 synthe	2.14e+02
439	6	1.2	382	2	S24282	E2 protein - human pa	2.14e+02	512	6	1.2	435	2	S19493	hypothetical protein	2.14e+02
440	6	1.2	382	1	S33573	nucleocapsid protein	2.14e+02	513	6	1.2	435	2	G71862	hypothetical protein	2.14e+02
441	6	1.2	382	2	S03936	ubiquitinol--cytochrome	2.14e+02	514	6	1.2	436	2	S42234	alpha-(1,3)-fucosyltr	2.14e+02
442	6	1.2	382	2	S47428	nucleocapsid protein	2.14e+02	515	6	1.2	440	1	B71293	paired box transcript	2.14e+02
443	6	1.2	382	1	VH1HPC	nucleocapsid protein	2.14e+02	516	6	1.2	441	2	J02191	hypothetical protein	2.14e+02
444	6	1.2	382	1	KIECGG	galactokinase (EC 2.7	2.14e+02	517	6	1.2	442	2	H71410	probable RNA helicase	2.14e+02
445	6	1.2	385	2	184552	prophage Sf6-like int	2.14e+02	518	6	1.2	442	2	S11712	transcription initiat	2.14e+02
446	6	1.2	386	2	H71701	succinyl-CoA synthea	2.14e+02	519	6	1.2	444	2	A48260	coriicolipin recept	2.14e+02
447	6	1.2	388	2	H65126	prophage Sf6-like int	2.14e+02	520	6	1.2	445	2	T01591	hypothetical protein	2.14e+02
448	6	1.2	388	2	A55597	oxytocin receptor - r	2.14e+02	521	6	1.2	448	2	A71547	hypothetical protein	2.14e+02
449	6	1.2	393	2	A55859	regulatory protein na	2.14e+02	522	6	1.2	448	1	A56018	transcription factor	2.14e+02
450	6	1.2	393	2	E64485	phosphoribosylglycna	2.14e+02	523	6	1.2	449	1	L1PG	triacylglycerol lipas	2.14e+02
451	6	1.2	393	2	G70891	hypothetical protein	2.14e+02	524	6	1.2	449	1	S30205	transcription factor	2.14e+02
452	6	1.2	395	2	A24709	ketol-acid reductoiso	2.14e+02	525	6	1.2	451	1	A40168	transcription factor	2.14e+02
453	6	1.2	395	1	J00396	nucleoside diphospho	2.14e+02	526	6	1.2	451	2	S49016	protein-tyrosine kin	2.14e+02
454	6	1.2	395	2	J05975	aurora-related kinase	2.14e+02	527	6	1.2	451	2	S58653	hypothetical protein	2.14e+02
455	6	1.2	395	2	Q0ECTR	hypothetical 45.2K pr	2.14e+02	528	6	1.2	454	2	I37257	coproporphyrinogen ox	2.14e+02
456	6	1.2	396	2	S76602	hypothetical protein	2.14e+02	529	6	1.2	454	2	B71914	alpha (1,3)-fucosyltr	2.14e+02
457	6	1.2	397	1	XNECY	tyrosine transaminase	2.14e+02	530	6	1.2	454	2	E70448	signal recognition pa	2.14e+02
458	6	1.2	397	2	S76786	hypothetical protein	2.14e+02	531	6	1.2	454	2	H70924	hypothetical protein	2.14e+02
459	6	1.2	399	2	I48911	interleukin-1 beta co	2.14e+02	532	6	1.2	460	1	A34180	sulfite oxidase (EC 1	2.14e+02
460	6	1.2	399	1	VCBBD	coat protein precurs	2.14e+02	533	6	1.2	461	2	S36593	E2 protein - human pa	2.14e+02
461	6	1.2	399	2	A46632	bombesin-like peptide	2.14e+02	534	6	1.2	463	2	S43539	argininosuccinate ly	2.14e+02

97	7	1.4	502	2	S36494	E2 protein - human pa	4.44e+00	170	6	1.2	114	2	S19433	hypothetical protein	2.14e+02
98	7	1.4	514	1	W2MLB5	E2 protein - human pa	4.44e+00	171	6	1.2	117	1	WMPSB	gene s protein - phag	2.14e+02
99	7	1.4	516	2	S44191	serine-type carboxype	4.44e+00	172	6	1.2	117	2	D69335	succinate dehydrogena	2.14e+02
100	7	1.4	524	2	I55595	splicing factor - hum	4.44e+00	173	6	1.2	118	2	S27476	hypothetical protein	2.14e+02
101	7	1.4	555	2	S69641	hypothetical protein	4.44e+00	174	6	1.2	118	2	S75320	hypothetical protein	2.14e+02
102	7	1.4	615	2	A53616	lamin B receptor - hu	4.44e+00	175	6	1.2	121	1	R5EC7	ribosomal protein L7/	2.14e+02
103	7	1.4	620	2	JC5567	lamin B receptor - ra	4.44e+00	176	6	1.2	121	1	R5PM2	ribosomal protein L7/	2.14e+02
104	7	1.4	620	2	F64408	coenzyme F420 hydroge	4.44e+00	177	6	1.2	125	1	F1SP4	photosystem I chain I	2.14e+02
105	7	1.4	634	2	A64521	outer membrane protei	4.44e+00	178	6	1.2	126	1	S16481	hypothetical protein	2.14e+02
106	7	1.4	637	2	A36427	Lamin B receptor - ch	4.44e+00	179	6	1.2	131	1	PC4345	phosphoinositide 3-Ki	2.14e+02
107	7	1.4	646	2	S47929	signal recognition pa	4.44e+00	180	6	1.2	131	1	BGB02	spermatid transiti	2.14e+02
108	7	1.4	656	2	S76505	hypothetical protein	4.44e+00	181	6	1.2	131	2	S33375	homeotic protein Abd-	2.14e+02
109	7	1.4	668	2	A71985	probable outer membra	4.44e+00	182	6	1.2	131	2	B61962	conserved hypotetica	2.14e+02
110	7	1.4	688	2	S61249	probable vitron prote	4.44e+00	183	6	1.2	133	2	S10038	hypothetical protein	2.14e+02
111	7	1.4	692	2	S37976	hypothetical protein	4.44e+00	184	6	1.2	133	2	S10038	hypothetical protein	2.14e+02
112	7	1.4	715	2	B33481	interferon-induced vi	4.44e+00	185	6	1.2	135	2	S69464	lactoyglutathione ly	2.14e+02
113	7	1.4	759	2	S50544	hypothetical protein	4.44e+00	186	6	1.2	135	2	S59464	hypothetical protein	2.14e+02
114	7	1.4	769	2	I56546	Shaw type potassium c	4.44e+00	187	6	1.2	136	2	I53300	interleukin-1-beta-co	2.14e+02
115	7	1.4	814	2	I40048	S-layer protein precu	4.44e+00	188	6	1.2	138	1	BGHM2	transformer-2 sex-det	2.14e+02
116	7	1.4	901	2	S50987	SOK1 protein - yeast	4.44e+00	189	6	1.2	139	2	A38612	spermatid transiti	2.14e+02
117	7	1.4	1032	2	A57514	RNA helicase HEL17 -	4.44e+00	190	6	1.2	139	2	D70756	insulin-like growth f	2.14e+02
118	7	1.4	1044	2	T02615	probable glycine dehy	4.44e+00	191	6	1.2	141	2	D71530	hypothetical protein	2.14e+02
119	7	1.4	1125	2	S67794	probable membrane pro	4.44e+00	192	6	1.2	141	2	A64662	probable L11 ribosoma	2.14e+02
120	7	1.4	1150	2	S49566	probable membrane pro	4.44e+00	193	6	1.2	144	2	JC2575	ATP synthase FO, subu	2.14e+02
121	7	1.4	1203	2	S26550	DNA-binding protein 5	4.44e+00	194	6	1.2	145	2	S53702	hypothetical 17K prot	2.14e+02
122	7	1.4	1220	2	A56236	probable RNA helicase	4.44e+00	195	6	1.2	146	2	S53702	hypothetical protein	2.14e+02
123	7	1.4	1444	2	S25327	cytoskeleton assembly	4.44e+00	196	6	1.2	148	2	F71367	probable PRS system	2.14e+02
124	7	1.4	1446	2	S60954	probable membrane pro	4.44e+00	197	6	1.2	152	2	S53640	protein kinase CLK3-1	2.14e+02
125	7	1.4	1457	2	T00385	probable serine/threo	4.44e+00	198	6	1.2	152	2	C71973	hypothetical protein	2.14e+02
126	7	1.4	1480	2	T00365	hypothetical protein	4.44e+00	199	6	1.2	153	2	JC5854	polyketide synthase (2.14e+02
127	7	1.4	1339	2	JC4387	epidermal growth fact	4.44e+00	200	6	1.2	155	2	A64533	conserved hypotetica	2.14e+02
128	7	1.4	1342	2	A36223	kinase-related transf	4.44e+00	201	6	1.2	156	2	A49547	nucleoside-diphosphat	2.14e+02
129	7	1.4	1342	2	I59164	HER3 protein precurs	4.44e+00	202	6	1.2	157	2	C70080	zinc finger (alternat	2.14e+02
130	7	1.4	1307	2	B47328	natural killer cell t	4.44e+00	203	6	1.2	158	2	S28744	conserved hypotetica	2.14e+02
131	7	1.4	1791	2	T02345	hypothetical protein	4.44e+00	204	6	1.2	161	2	S01149	NAH dehydrogenase (u	2.14e+02
132	7	1.4	1926	2	JC4842	DNA-binding nuclear P	4.44e+00	205	6	1.2	161	2	C70613	H+-transporting ATP s	2.14e+02
133	7	1.4	2161	2	JH0564	calcium channel alpha	4.44e+00	206	6	1.2	163	2	I38191	probable sece - Mycob	2.14e+02
134	7	1.4	2181	2	A38198	calcium channel alpha	4.44e+00	207	6	1.2	163	2	S61520	nucleic acid binding	2.14e+02
135	7	1.4	2278	2	S56274	FAB1 protein - yeast	4.44e+00	208	6	1.2	163	2	S14572	hemoglobin alpha chai	2.14e+02
136	7	1.4	4128	2	JC6306	protein kinase (EC 2.	4.44e+00	209	6	1.2	170	2	S05303	hemoglobin protein A	2.14e+02
137	6	1.2	14	2	B56884	lamin B receptor - tu	2.14e+02	210	6	1.2	173	2	E69183	placoglycin precurs	2.14e+02
138	6	1.2	20	2	S28405	protein kinase - EC 2.	2.14e+02	211	6	1.2	175	2	S36749	hypothetical protein	2.14e+02
139	6	1.2	31	2	A41269	lamin B receptor - tu	2.14e+02	212	6	1.2	177	2	S39519	transcription factor	2.14e+02
140	6	1.2	35	2	B45316	sperm-specific protam	2.14e+02	213	6	1.2	178	2	B69844	H+-transporting ATP s	2.14e+02
141	6	1.2	36	2	A37172	collagen alpha 1(XII)	2.14e+02	214	6	1.2	178	2	S26044	hypothetical protein	2.14e+02
142	6	1.2	37	2	S29829	protamine Z3 - smalle	2.14e+02	215	6	1.2	180	2	H64333	sex-determining prote	2.14e+02
143	6	1.2	40	2	B27740	protamine Z3 - smalle	2.14e+02	216	6	1.2	180	2	S56056	hypothetical protein	2.14e+02
144	6	1.2	41	2	I68638	gas-vesicle protein -	2.14e+02	217	6	1.2	180	2	E70145	ribosomal protein L18	2.14e+02
145	6	1.2	45	2	G02087	retinoblastoma suscep	2.14e+02	218	6	1.2	181	2	A40607	hypothetical protein	2.14e+02
146	6	1.2	55	1	VHNVAC	RNA binding motif pro	2.14e+02	219	6	1.2	183	2	E70588	monofunctional chori	2.14e+02
147	6	1.2	69	1	VHNVBM	nucleocapsid protein	2.14e+02	220	6	1.2	184	2	S46808	hypothetical protein	2.14e+02
148	6	1.2	65	1	S39424	protamine P1 - Austa	2.14e+02	221	6	1.2	184	2	S26046	sex-determining prote	2.14e+02
149	6	1.2	76	2	S28127	gas-vesicle protein 9	2.14e+02	222	6	1.2	185	2	F71349	probable transcriptio	2.14e+02
150	6	1.2	76	2	S01423	gas-vesicle protein -	2.14e+02	223	6	1.2	186	2	A70109	hypothetical protein	2.14e+02
151	6	1.2	87	2	A55571	chloride conductance	2.14e+02	224	6	1.2	186	2	S74835	hypothetical protein	2.14e+02
152	6	1.2	88	2	A70946	probable ribosomal pr	2.14e+02	225	6	1.2	187	1	Q0EC30	prelamin peptidase de	2.14e+02
153	6	1.2	83	2	G64756	YK1 protein precurs	2.14e+02	226	6	1.2	188	1	R5R781	ribosomal protein L18	2.14e+02
154	6	1.2	84	2	D28928	pregnancy-specific be	2.14e+02	227	6	1.2	188	2	S38352	ribosomal protein L18	2.14e+02
155	6	1.2	85	2	S39512	DNA-directed RNA poly	2.14e+02	228	6	1.2	189	2	I65744	gene MAC30 protein -	2.14e+02
156	6	1.2	87	2	S07323	gas-vesicle protein -	2.14e+02	229	6	1.2	192	2	E71081	hypothetical protein	2.14e+02
157	6	1.2	91	2	S59647	hypothetical protein	2.14e+02	230	6	1.2	192	2	A64377	hypothetical protein	2.14e+02
158	6	1.2	92	2	S34115	sperm-specific protei	2.14e+02	231	6	1.2	195	2	B36870	mannitol operon rep	2.14e+02
159	6	1.2	95	2	A56644	inverted repeat compo	2.14e+02	232	6	1.2	196	2	D64918	glucuronide repressor	2.14e+02
160	6	1.2	99	1	CUSU	plastocyanin - shepe	2.14e+02	233	6	1.2	197	2	A29648	female-specific trans	2.14e+02
161	6	1.2	99	1	CUPD	tyrosin inhibitor 2 p	2.14e+02	234	6	1.2	199	2	JU0052	coat protein - cucurb	2.14e+02
162	6	1.2	99	2	S65661	hypothetical protein	2.14e+02	235	6	1.2	200	1	R3MD4	puromycin N-acetylra	2.14e+02
163	6	1.2	100	2	S50733	hypothetical protein	2.14e+02	236	6	1.2	201	2	JU0049	ribosomal protein S4,	2.14e+02
164	6	1.2	101	2	D71426	hypothetical protein	2.14e+02	237	6	1.2	203	2	S60449	proteinase inhibitor	2.14e+02
165	6	1.2	101	2	S51384	hypothetical protein	2.14e+02	238	6	1.2	203	2	S60449	probable membrane pro	2.14e+02
166	6	1.2	101	2	D69359	H+-transporting ATP s	2.14e+02	239	6	1.2	204	2	T02386	hypothetical protein	2.14e+02
167	6	1.2	103	2	D70159	ribosomal protein S10	2.14e+02	240	6	1.2	205	2	S19634	nucleolar protein GAR	2.14e+02
168	6	1.2	104	1	BVECAA	arct protein - Escher	2.14e+02	241	6	1.2	209	1	S35108	cell division protein	2.14e+02
169	6	1.2	106	2	S32964	hypothetical protein	2.14e+02	242	6	1.2	209	1	S48459	probable dual specifi	2.14e+02

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPsrch_gp  protein - protein database search, using Smith-Waterman algorithm

Run on:      Tue Aug 10 11:42:14 1999;      MasPar time 21.14 Seconds
Tabular output not generated.      953.603 Million cell updates/sec

```

Title:	>US-08-287-669-19
Description:	(1-503) from US08287669.pep
Perfect Score:	503
Sequence:	1 MMRDRLSSLEIRINIMFSSH.....MTSLKKFTYFPPARSAV 503

Scoring table: TABLE unitprotable
Gap 60

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 1000 summaries

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 3.750; Variance 0.445; scale 8.419

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	452	89.9	503	2	A49429	interleukin-1 beta-co	0.00e+00
2	8	1.6	138	2	I48723	ciliary neurotrophic	5.18e-02
3	8	1.6	159	1	UNRBCF	ciliary neurotrophic	5.18e-02
4	8	1.6	200	1	UNRBCF	ciliary neurotrophic	5.18e-02
5	8	1.6	284	2	JC6125	U2 small nuclear ribo	5.18e-02
6	8	1.6	301	2	A70039	ABC transporter (ATP-	5.18e-02
7	8	1.6	485	2	S75895	hypothetical protein	5.18e-02
8	8	1.6	506	1	W2W447	E2 protein - human pa	5.18e-02
9	8	1.6	821	2	S359983	eps8 protein - mouse	5.18e-02
10	8	1.6	822	2	I387228	epidermal growth fact	5.18e-02
11	7	1.4	34	2	S21080	sperm protein EMI - m	4.44e+00
12	7	1.4	56	2	CS58213	prolamine II - Americ	4.44e+00
13	7	1.4	58	2	S34045	prolamine - North Ame	4.44e+00
14	7	1.4	58	2	A58208	prolamine T-1 - paint	4.44e+00
15	7	1.4	61	2	S39425	prolamine I - duckb1	4.44e+00
16	7	1.4	62	2	B58123	prolamine I - America	4.44e+00
17	7	1.4	101	2	S65494	sperm-specific protei	4.44e+00
18	7	1.4	135	2	A36828	or10 protein - Atog	4.44e+00
19	7	1.4	147	2	I51168	hemoglobin beta-4b ch	4.44e+00
20	7	1.4	148	2	A45316	sperm-specific histon	4.44e+00
21	7	1.4	158	3	B45319	PL-III' and PL-IV comm	4.44e+00
22	7	1.4	162	1	Q58E19	BKRF2 protein - human	4.44e+00
23	7	1.4	164	2	I45016	X16 protein - mouse	4.44e+00

24	1.4	164	2	154089	pre-mRNA splicing fac
25	1.4	175	2	A39111	T-cell surface glycop
26	1.4	198	2	S53911	ribosomal protein L16
27	1.4	199	2	S26045	transformer sex-deter
28	1.4	201	2	S26045	sex-determining prote
29	1.4	202	2	A42410	ryudocan precursor -
30	1.4	203	2	S63633	hypothetical protein
31	1.4	203	3	A45317	PL-II* and PL-IV comm
32	1.4	204	2	S35265	actin (clone Tac)
33	1.4	204	2	F36950	urease accessory prot
34	1.4	218	2	E70645	probable fucosylase-1
35	1.4	221	2	B42701	PR264 protein - chick
36	1.4	221	2	A42634	splicing factor SC35
37	1.4	221	2	A42701	PR264/SC35 protein -
38	1.4	238	2	A57198	splicing factor, araf
39	1.4	240	2	A46119	U2 snRNP auxiliary fa
40	1.4	244	2	A44822	F44E2.3 protein - Cae
41	1.4	248	2	A40040	alternative splicing
42	1.4	252	2	C43679	ORF3 protein - Autogr
43	1.4	253	2	E69098	phosphate transport S
44	1.4	268	2	S61003	hypothetical protein
45	1.4	269	2	B47112	growth response prote
46	1.4	272	2	S59042	splicing factor SRP40
47	1.4	274	2	A55335	myelin regulatory fac
48	1.4	288	2	S68798	RNA-binding protein S
49	1.4	299	2	AHRB	Ig alpha chain C regi
50	1.4	303	2	S71185	splicing factor SF-2
51	1.4	310	2	F64146	hypothetical 35.2 kD
52	1.4	315	2	C64982	hypothetical 35.2 kD
53	1.4	315	1	MNXR4S	nonstructural protein
54	1.4	334	1	M2WLB4	E2 protein - bovine p
55	1.4	335	2	S09275	Ig alpha chain C regi
56	1.4	338	2	S09276	Ig alpha chain C regi
57	1.4	339	2	S09264	Ig alpha chain C regi
58	1.4	343	2	S09272	Ig alpha chain C regi
59	1.4	344	2	S59043	splicing factor SRP55
60	1.4	346	2	S48380	IMP3 protein - yeast
61	1.4	348	2	S09283	Ig alpha chain C regi
62	1.4	348	2	S09270	Ig alpha chain C regi
63	1.4	350	2	A40459	nuclear phosphoprotei
64	1.4	352	2	S09266	Ig alpha chain C regi
65	1.4	357	2	S09269	Ig alpha chain C regi
66	1.4	357	2	S09265	Ig alpha chain C regi
67	1.4	357	2	S09267	Ig alpha chain C regi
68	1.4	358	2	S09268	Ig alpha chain C regi
69	1.4	360	2	S09271	Ig alpha chain C regi
70	1.4	364	2	S62183	cysteine proteinase T
71	1.4	373	2	S47911	PSI protein - yeast
72	1.4	374	2	A37282	52K active chromatin
73	1.4	376	1	A48681	platelet thrombin in
74	1.4	378	2	A57488	proteinase inhibitor
75	1.4	379	2	S49959	40K protein - frog vi
76	1.4	379	2	S55421	acyl-CoA dehydrogena
77	1.4	396	2	S58223	LSRI protein - yeast
78	1.4	410	2	S18157	gluculin 2 - eastern
79	1.4	418	2	B57511	interleukin-1 beta co
80	1.4	429	2	D70468	conserved hypothetical
81	1.4	431	2	I76773	ATP binding component
82	1.4	438	2	B69023	conserved hypothetical
83	1.4	440	2	E70326	conserved hypothetical
84	1.4	448	2	A36311	70K UI small nuclear
85	1.4	452	2	S36482	E2 protein - human pa
86	1.4	456	2	S61558	hypothetical protein
87	1.4	477	2	S22027	paramyosin, mini - fr
88	1.4	478	1	UNPDS9	aspartate ammonia-ly
89	1.4	483	2	FN0099	son3 protein - human
90	1.4	483	2	S36470	E2 protein - human pa
91	1.4	484	2	A40988	54K arginine-rich nuc
92	1.4	488	2	S36570	E2 protein - human pa
93	1.4	488	2	S18156	globulin 1 - eastern
94	1.4	493	2	S36488	E2 protein - human pa
95	1.4	494	2	A48133	pre-mRNA splicing SFP
96	1.4	498	1	W2WL80	E2 protein - human pa


```
Db 1 mmrgdrrslernmmfshlkvdeilevliakvlnsdngmlnscgtvrekrrreivka 60
   |||||||
Qy 1 MRQDRRSLEERNIMMFSSHLKVEILEVLIAQVLNSDNGMINSCTVREKREIYKA 60
   |||||||
Db 61 vgrtdvafdafydgdlrrstghegleavleplarsvdsnavefecpmshrrsralspa 120
   |||||||
Qy 61 VQRSDVAADAFYDALRSTGHEGLEAVLEPLARSVDNNAVEFECPMSPASHRRSRALS PA 120
   |||||||
Db 121 gylspclrvhrdsavssvftsdyqdiysrararsrsralhsdqrhnyssppynaafpsqps 180
   |||||||
Qy 121 GYTSPTRVHRDSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSPPYNAAFPSQPS 180
   |||||||
Db 181 anssftgcsslgysssrnrsfskaspqaylfheedmnfvadaptisryfdektmyrnfes 240
   |||||||
Qy 181 ANSSFTGCSLSGSSSRNRSFSKASGPTQYIFHEEDMNFVDAPISRYFDEKTYRNFS 240
   |||||||
Db 241 prgmexlnehefegmptnrgtkadkdnltlfrcmgytvlckdltgrmlltirdfak 300
   |||||||
Qy 241 PRGMCLINNEHEFEGMPTNRGTRKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
   |||||||
Db 301 heshgdsallvllshgeenvllgvddipistheiydlinaanaprlnkpkivfqacrg 360
   |||||||
Qy 301 HESHGDSALLVLLSHGEENVLLGVDDIPISTHEIYDLINAAANAPRLANKPKIVFQACRG 360
   |||||||
Db 361 errdngfpvldsdyvpaflrrgwdnrdgplfnfigcvrpgyqvwrkksqadiliry 420
   |||||||
Qy 361 ERRDNGFPVLDSYDVPAFLRRGWDNRDGPLFNFIGCVRPGYQVWRKKPSQADILIRYA 420
   |||||||
Db 421 ttagyvswnsarqswfigavcevfsthakmdvvelillevnkka 466
   |||||||
Qy 421 TTAGYVSWNSARQSWFIGAVCEVFSTHAKMDVVELLLEVANKKA 466
   |||||||
```

Search completed: Tue Aug 10 11:46:04 1999
Job time : 117 secs.

Db 421 ttaqyvswnsarsgswfifgavcevfsthakmdvnulltevnkva 466
 |||||||
 421 TTAQYVSWNSARSGSWFIFGAVCEVFSTHAKMDVNULLTEVNKVA 466

RESULT 14

ID R45303 standard; Protein: 503 AA.

AC R45303;

DE 01-JUL-1994 (first entry)

DE Ced-3 mutant G316.

KW C.elegans; cell death; gene: ced-3; inhibition: human; parasite;

KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;

KW inflammatory response; nematode; diagnosis; myocardial infarction;

KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;

KW pathogenic infection; hair loss; cancer; autoreactive antibody.

OS Caenorhabditis elegans.

EH Key Location/Qualifiers

FT misc_difference 316

FT /Label= G316

PN W09325694-A.

PD 23-DEC-1993.

PF 14-JUN-1993; U05705.

PR 12-JUN-1992; US-897788.

PR 20-NOV-1992; US-984182.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Horvitz HR, Shahan S, Yuan J;

PI WPI; 94-007551/01.

PT Agents which affect activity of cell death genes - used to

PT develop drugs for treating conditions characterised by cell death

PT or proliferation

PS Claim 99; : 132pp; English.

FT The sequences given in R45276-328 indicate the positions of possible

FT mutations of the C.elegans cell death gene, ced-3. These mutations

FT occur at conserved amino acid residues of the Ced-3 protein.

CC Fragments of the amino terminal of these proteins act as inhibitors of

CC Ced-3. The ced-3 gene has considerable similarity to human interleukin

CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the

CC active cytokine and is involved in inflammatory response in humans.

CC The similarity between the two sequences suggests that inhibitors of

CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3

CC proteins have an overall amino acid similarity of 28%. The ced-3

CC inhibitors may be used for identifying agents which affect the activity

CC of a gene belonging to the to the ced-3/ICE family of genes and for

CC diagnosis of diseases characterised by cell death. They can also be

CC used to develop drugs for treating conditions characterised by cell

CC deaths such as myocardial infarction, stroke, degenerative disease,

CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,

CC or drugs for reducing the proliferative capacity or size of a

CC population of cells such as cancerous cells, cells which produce

CC autoreactive antibodies, infected cells, hair follicle cells or cells

CC which are critical to the life of a parasite, pest or recombinant

CC organism. They may also be used in the diagnosis of inflammatory

CC disease.

CC Sequence 503 AA;

SO

Query Match 92.4%; Score 465; DB 9; Length 503;

Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrrslernimfshlkvdeilevliakvlnsdngdmniscgvrekrreivka 60

1 |||||||

Qy 1 MRRORRSLERNIMFSSHLKVDLEILVLIKQVLDNDGDMNISCCTVREKRREIVKA 60

1 |||||||

Db 61 vgrivdafaadaiiristgheglavleplarsvdsnavefecpmaspirralspa 120

1 |||||||

Qy 61 VORRDVAFADAIYDALKSTGHEGLAVLEPLARSDVNAVEFECPMASPIRRALSPA 120

1 |||||||

Db 121 gytsptrvrhdsrvssvftsdyqdlrsarsrsrrahsdrrhnyssppvnafpssps 180

1 |||||||

Qy 121 GYTSPTVRHDSRVSSVFTSYQDLRSARSRRAHSSDRHNYSSPPVNAFPSSPS 180

1 |||||||

Db 181 anssftgcsaljsssrnsfksaegptqyifhnedmfvdaptisrvfdektmyrnfsk 240

1 |||||||

Qy 181 ANSFTGCSALJSSSRNSRNSFSKASGPTQYIFHEEDMNFVDAPTISRVEDKTYRNFS 240
 |||||||
 Db 241 prgmclltnnehfegmptnrgtkadnltlnlfrcmgyvicknltgrgmlltrdfak 300
 |||||||
 Qy 241 PRGMCLITNNEHFEGMPTNRTGKADNLTNLFRCMGYVICKNLTGRGMILLTRDFAK 300
 |||||||
 Db 301 heshgdsallvllshxeenvlsgvdipistheiydlinaanaprlanpkivfagacg 360
 |||||||
 Qy 301 HESHGDSALLVLLSHXEENVLSGVDIPISTHEITDILMAANAPRLANKPIVFOACRG 360
 |||||||
 Db 361 errdngfpvldsvdygpaflirgwdndpplnfigvrvpqvqwrkksqadilliya 420
 |||||||
 Qy 361 ERRDNGFPVLDSVDYGPAPFLIRGWDNDPPLNFIGVCVPQVQWRKKPSQADILIRYA 420
 |||||||
 Db 421 ttaqyvswnsarsgswfifgavcevfsthakmdvnulltevnkva 466
 |||||||
 Qy 421 TTAQYVSWNSARSGSWFIFGAVCEVFSTHAKMDVNULLTEVNKVA 466
 |||||||

RESULT 15

ID R45279 standard; Protein: 503 AA.

AC R45279;

DE 01-JUL-1994 (first entry)

DE Ced-3 mutant L246.

KW C.elegans; cell death; gene: ced-3; inhibition: human; parasite;

KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;

KW inflammatory response; nematode; diagnosis; myocardial infarction;

KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;

KW pathogenic infection; hair loss; cancer; autoreactive antibody.

OS Caenorhabditis elegans.

EH Key Location/Qualifiers

FT misc_difference 246

FT /Label= L246

PN W09325694-A.

PD 23-DEC-1993.

PF 14-JUN-1993; U05705.

PR 12-JUN-1992; US-897788.

PR 20-NOV-1992; US-984182.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Horvitz HR, Shahan S, Yuan J;

PI WPI; 94-007551/01.

PT Agents which affect activity of cell death genes - used to

PT develop drugs for treating conditions characterised by cell death

PT or proliferation

PS Claim 99; : 132pp; English.

FT The sequences given in R45276-328 indicate the positions of possible

FT mutations of the C.elegans cell death gene, ced-3. These mutations

FT occur at conserved amino acid residues of the Ced-3 protein.

CC Fragments of the amino terminal of these proteins act as inhibitors of

CC Ced-3. The ced-3 gene has considerable similarity to human interleukin

CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the

CC active cytokine and is involved in inflammatory response in humans.

CC The similarity between the two sequences suggests that inhibitors of

CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3

CC proteins have an overall amino acid similarity of 28%. The ced-3

CC inhibitors may be used for identifying agents which affect the activity

CC of a gene belonging to the to the ced-3/ICE family of genes and for

CC diagnosis of diseases characterised by cell death. They can also be

CC used to develop drugs for treating conditions characterised by cell

CC deaths such as myocardial infarction, stroke, degenerative disease,

CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,

CC or drugs for reducing the proliferative capacity or size of a

CC population of cells such as cancerous cells, cells which produce

CC autoreactive antibodies, infected cells, hair follicle cells or cells

CC which are critical to the life of a parasite, pest or recombinant

CC organism. They may also be used in the diagnosis of inflammatory

CC disease.

CC Sequence 503 AA;

SO

Query Match 92.4%; Score 465; DB 9; Length 503;

Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 99; : 132p; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -beta convertase (ICE), which converts pro-interleukin-beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 SQ Sequence 503 AA:
 Query Match 92.4%; Score 465; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 1 mmqdrtsllernimfshlkvdeillevlakvlnsdngdmnscgtvrekrrtelyka 60
 QY 1 MMRDRRSLLERNIMFSSHLKVDLEILLAKVQLNSDNGDMNISCCTVREKRRLEYKA 60
 Db 61 vqrrgdaafadafydaalrstqheglaevleplarsvdsnavfeecmpasprsralspa 120
 QY 61 VQRGDVAFAFADAFYDALRSTQHEGLAEVLEPLARSDVSNAAVEFECMPASRRSRALSPA 120
 Db 121 gytsptrvhndsvsvsftsyqdiysrarstrsralshsdtrhysppnaifspqss 180
 QY 121 GYTSPTRVHNRDSVSVSFTSYQDIYSRARSTRSRALHSSDRHNYSPPVNAFPSPQSS 180
 Db 181 anssftgcsslgysssrnsrfskasgptqyifhnedmfvdaptisrvfdektmyrnfs 240
 QY 181 ANSFTGCSSLGYSSSRNSRFSKASGPTQYIFHEEDMNFVDAPTRISRVFEKMYRNFS 240
 Db 241 prgmcllxxnehfegmpttrngtkadkdnlnlfrcmgytvickdnltgrcmlltirdfak 300
 QY 241 PRGMCLLXNEHFEGMPTTRNGTKADKDNLNLFRCMGTYTICKDNLTRGMLLTIRDFAK 300
 Db 301 heshgdsaalvllshgeenvllgyvddipistheylldlanaaprjanpkxlvfyagcrg 360
 QY 301 HESHGDSALVLLSHGEEENVLLIGVDDIPISTHEIYDLLNANAPRLANKPKXIVFYVQACRG 360
 Db 361 erlingfpvldsvdgvpaflrrgwdndgrplfnlfgcvrpgvqvwkxkpsqadllirya 420
 QY 361 ERRONGFPVLDSDVGPAPFLRRGWDNDGRPLFNFGCVRPQVOQVWKRKXPSQADILIRYA 420
 Db 421 ttaayvswrnasrgswfagavcevfshakmdmveilleenkkva 466
 QY 421 TTAQVSWRNSARSGSWFIQAVCEVFSTHAKMDVLELLEENKKVA 466
 RESULT 13
 TD R45294 standard; Protein: 503 AA.
 AC R45294;
 DT 01-JUL-1994 (first entry)
 DE Ced-3 mutant M291.
 KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
 KW interleukin-beta convertase; ICE; pro-interleukin-beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;

KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH key Location/Qualifiers
 FT misc_difference 291
 FT /label= M291
 PN WO9325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05705
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shalem S, Yuan J;
 DR WPI: 94-007551/01.
 PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 99; : 132p; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -beta convertase (ICE), which converts pro-interleukin-beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 SQ Sequence 503 AA:
 Query Match 92.4%; Score 465; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 1 mmqdrtsllernimfshlkvdeillevlakvlnsdngdmnscgtvrekrrtelyka 60
 QY 1 MMRDRRSLLERNIMFSSHLKVDLEILLAKVQLNSDNGDMNISCCTVREKRRLEYKA 60
 Db 61 vqrrgdaafadafydaalrstqheglaevleplarsvdsnavfeecmpasprsralspa 120
 QY 61 VQRGDVAFAFADAFYDALRSTQHEGLAEVLEPLARSDVSNAAVEFECMPASRRSRALSPA 120
 Db 121 gytsptrvhndsvsvsftsyqdiysrarstrsralshsdtrhysppnaifspqss 180
 QY 121 GYTSPTRVHNRDSVSVSFTSYQDIYSRARSTRSRALHSSDRHNYSPPVNAFPSPQSS 180
 Db 181 anssftgcsslgysssrnsrfskasgptqyifhnedmfvdaptisrvfdektmyrnfs 240
 QY 181 ANSFTGCSSLGYSSSRNSRFSKASGPTQYIFHEEDMNFVDAPTRISRVFEKMYRNFS 240
 Db 241 prgmcllxxnehfegmpttrngtkadkdnlnlfrcmgytvickdnltgrcmlltirdfak 300
 QY 241 PRGMCLLXNEHFEGMPTTRNGTKADKDNLNLFRCMGTYTICKDNLTRGMLLTIRDFAK 300
 Db 301 heshgdsaalvllshgeenvllgyvddipistheylldlanaaprjanpkxlvfyagcrg 360
 QY 301 HESHGDSALVLLSHGEEENVLLIGVDDIPISTHEIYDLLNANAPRLANKPKXIVFYVQACRG 360
 Db 361 erlingfpvldsvdgvpaflrrgwdndgrplfnlfgcvrpgvqvwkxkpsqadllirya 420
 QY 361 ERRONGFPVLDSDVGPAPFLRRGWDNDGRPLFNFGCVRPQVOQVWKRKXPSQADILIRYA 420

PN W09325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death
PT or proliferation
CS Claim 99; : 132pp; English.
CC The sequences given in R45376-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -beta convertase (ICE), which converts pro-interleukin-beta to the
CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
CC proteins have an overall amino acid similarity of 28%. The ced-3
CC inhibitors may be used for identifying agents which affect the activity

RESULT	12
ID	R45281 standard; Protein; 503 AA.
AC	R45281;
DT	01-JUL-1994 (first entry)
DE	Ced-3 mutant I248
KW	C.elegans; cell death; gene; ced-3; inhibition; human; par
KW	interleukin-1beta convertase; ICE; pro-interleukin-1beta;
KW	inflammatory response; nematode; diagnosis; myocardial inf
KW	stroke; degenerative disease; traumatic brain injury; hypo
KW	pathogenic infection; hair loss; cancer; autoreactive anti
OS	Caenorhabditis elegans.
FH	Key
FT	misc_difference 248
FT	location/Qualifiers
FN	
PD	W09325694-A.
PD	23-DEC-1993.
PF	14-JUN-1993; U05705.
PR	12-JUN-1992; US-897788.
PR	20-NOV-1992; US-984182.
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.
PI	Horvitz BR, Shaham S, Yuan J;
DT	WPI; 94-007551/01.
DR	Agents which affect activity of cell death genes - used to

|||||
QY 361 ERDNGEFPVLDSVGVPAFLRGMNDNDGFLNFGCVRPQVOQWRKKPSQADILIRYA 420
Db 421 ttaqyvwrsnarsgswfigavcevfsthakmdvveltlewnkva 466
QY 421 TTAQYVSWRSNARSWSFIQAVCEVFSTHAKMDVVELLEVNKKVA 466
RESULT 9
ID R45326 standard; Protein: 503 AA.
AC R45326;
DT 01-JUL-1994 (first entry)
DE ced-3 mutant L488
KW C.elegans; cell death; gene: ced-3; inhibition; human; parasite;
interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
L1 misc_difference 488 /label= L488
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; 005705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI: 94-007551/01.
PT Agents which affect activity of cell death genes - used to
develop drugs for treating conditions characterised by cell death
or proliferation
PS Claim 99: 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
mutations of the C.elegans cell death gene, ced-3. These mutations
occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
Ced-3. The ced-3 gene has considerable similarity to human interleukin
-1beta convertase (ICE), which converts pro-interleukin-1beta to the
active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
proteins have an overall amino acid similarity of 28%. The ced-3
inhibitors may be used for identifying agents which affect the activity
of a gene belonging to the to the ced-3/ICE family of genes and for
diagnosis of diseases characterised by cell death. They can also be
used to develop drugs for treating conditions characterised by cell
deaths such as myocardial infarction, stroke, degenerative disease,
traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
or drugs for reducing the proliferative capacity or size of a
population of cells such as cancerous cells, cells which produce
autoreactive antibodies, infected cells, hair follicle cells or cells
which are critical to the life of a parasite, pest or recombinant
organism. They may also be used in the diagnosis of inflammatory
disease.
CC disease.
SQ Sequence 503 AA:
Query Match 92.6%; Score 466; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 181 ansstfgcsslgysssnrnsfskaspctqyifhneadmfvadptisrfvfeakmyrnfs 240
QY 181 ANSSFTFGSSSLGYSSSNRNSFSKASCPLOYIFHEEDNMFVADPTISRVEFEKMYRNFS 240
Db 241 prngcllmmhbfqmptrngtkadknlntfcmgytvicldnltgrmllttrdfak 300
QY 241 PRNGCLLMMHBFQMPTRNGTKADKDLNLTFCMGYTVICLDNLTGRMLLTTRDFAK 300
Db 301 heshgdsailvllshgeenvliigvddlpiethetldlNANAPRLANKRKiFVQACRG 360
QY 301 HESHODSAILVILSHGEENVLIIGVDDLPISHTHEITDLDNANAPRLANKRKIFVQACRG 360
Db 361 erdnqfvlsvdgvpaflrgrvndnrgplfnflgcvrpgvqvwrkpsqadiliry 420
QY 361 ERDNGEFPVLDSVGVPAFLRGMNDNDGFLNFGCVRPQVOQWRKKPSQADILIRYA 420
Db 421 ttaqyvwrsnarsgswfigavcevfsthakmdvveltlewnkva 466
QY 421 TTAQYVSWRSNARSWSFIQAVCEVFSTHAKMDVVELLEVNKKVA 466
RESULT 10
ID R45282 standard; Protein: 503 AA.
AC R45282;
DT 01-JUL-1994 (first entry)
DE Ced-3 mutant N250.
KW C.elegans; cell death; gene: ced-3; inhibition; human; parasite;
interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
L1 misc_difference 250 /label= N250
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; 005705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI: 94-007551/01.
PT Agents which affect activity of cell death genes - used to
develop drugs for treating conditions characterised by cell death
or proliferation
PS Claim 99: 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
mutations of the C.elegans cell death gene, ced-3. These mutations
occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
Ced-3. The ced-3 gene has considerable similarity to human interleukin
-1beta convertase (ICE), which converts pro-interleukin-1beta to the
active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
proteins have an overall amino acid similarity of 28%. The ced-3
inhibitors may be used for identifying agents which affect the activity
of a gene belonging to the to the ced-3/ICE family of genes and for
diagnosis of diseases characterised by cell death. They can also be
used to develop drugs for treating conditions characterised by cell
deaths such as myocardial infarction, stroke, degenerative disease,
traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
or drugs for reducing the proliferative capacity or size of a
population of cells such as cancerous cells, cells which produce
autoreactive antibodies, infected cells, hair follicle cells or cells
which are critical to the life of a parasite, pest or recombinant
organism. They may also be used in the diagnosis of inflammatory
disease.
CC disease.
SQ Sequence 503 AA:
Query Match 92.4%; Score 465; DB 9; Length 503;

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI; 94-007551/01.
 DR N-PSDB; Q54401.
 PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 2; Fig 3; 132pp; English.
 CC This sequence is encoded by the C.elegans cell death gene, ced-3.
 CC Fragments of the amino terminal of this protein act as inhibitors of
 CC ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 CC Sequence 503 AA:

Query Match 92.6%; Score 466; DB 9; Length 503;

Best Local Similarity 100.0%; Pred. No. 0.00e+00; Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrgdrsllelnimfshlkvdellevlajakvlnsdngmnsqgvrekrrivya 60
 1 MMRODRSLLELNIMFSSHLKVDLEVLIAKQVLSNDGMINSCGVREKRRIVYA 60
 QY 1 MMRODRSLLELNIMFSSHLKVDLEVLIAKQVLSNDGMINSCGVREKRRIVYA 60
 Db 61 vqrgvafafafdalrstrgheglaeavlleplarsvsnavefecmpspashrralspa 120
 61 VQRGDVAFDAFALRSTRGHEGLAEVLEPLARSDVSNAVEFCMPSPASHRRALSPA 120
 QY 61 VQRGDVAFDAFALRSTRGHEGLAEVLEPLARSDVSNAVEFCMPSPASHRRALSPA 120
 Db 121 gytsptrvhdsvsvsfstsygdlysrarsrsralhsdthmssppvnaafpsqps 180
 121 GYTSPTRVHDSSVSFSSTSYGDLYSRARSRSRALHSDDHNSPPVNAFPSPQPS 180
 QY 121 GYTSPTRVHDSSVSFSSTSYGDLYSRARSRSRALHSDDHNSPPVNAFPSPQPS 180
 Db 181 ansstfgcsllgysssrnsfskasgptqyifheedmfvdapltisrvfdektmyrnfs 240
 181 ANSSTFGCSLLGYSSSRNSFSKASGPTQYIFHEEDMFVDAPLTISRVEDEKTMYNFS 240
 QY 181 ANSSTFGCSLLGYSSSRNSFSKASGPTQYIFHEEDMFVDAPLTISRVEDEKTMYNFS 240
 Db 241 prgmcllinnehfeqmptrngtkadkdnltlfrcmgyvickdnltgrgmlltridfak 300
 241 PRGMCLLINNEHEQMPTRNGTKADKDNLTNLFRCMGYVICKDNLTRGMLLTRIDFAK 300
 QY 241 PRGMCLLINNEHEQMPTRNGTKADKDNLTNLFRCMGYVICKDNLTRGMLLTRIDFAK 300
 Db 301 heshgsallvllshgeenvllgvdipistheilydllaanaaprlanpkivfoacrg 360
 301 HESHGSALLVLLSHGEENVLLGVDDIPISTHEIYDLLAANAAPRLANPKIVFOACRG 360
 QY 301 HESHGSALLVLLSHGEENVLLGVDDIPISTHEIYDLLAANAAPRLANPKIVFOACRG 360
 Db 361 errdngfprlvdsvdgyvafirgwdnrdgplfnflgcvrpgvqvwirkpsqadlilrya 420
 361 ERRDNGFPRVLDSVDGYVAFIRGWDNRDGPLFNFLGCVRPQVQVWIRKPPSQADLILRYA 420
 QY 361 ERRDNGFPRVLDSVDGYVAFIRGWDNRDGPLFNFLGCVRPQVQVWIRKPPSQADLILRYA 420
 Db 421 ttaqyvswnsargswfivavcevfisthakndvvevlltevnkva 466
 421 TTAQYVSWNSARGSWFIVAVCEVFISTHAKNDVVEVLLTEVNKVA 466
 QY 421 TTAQYVSWNSARGSWFIVAVCEVFISTHAKNDVVEVLLTEVNKVA 466

RESULT 8

ID R45328 standard; Protein; 503 AA.
 AC R45328;
 DT 01-JUL-1994 (first entry)
 DE Ced-3 mutant P496.
 KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;

KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT misc_difference 496
 FT /label= P496
 PN W09325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI; 94-007551/01.
 PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 99; 132pp; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 CC Sequence 503 AA:

Query Match 92.6%; Score 466; DB 9; Length 503;

Best Local Similarity 100.0%; Pred. No. 0.00e+00; Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrgdrsllelnimfshlkvdellevlajakvlnsdngmnsqgvrekrrivya 60
 1 MMRODRSLLELNIMFSSHLKVDLEVLIAKQVLSNDGMINSCGVREKRRIVYA 60
 QY 1 MMRODRSLLELNIMFSSHLKVDLEVLIAKQVLSNDGMINSCGVREKRRIVYA 60
 Db 61 vqrgvafafafdalrstrgheglaeavlleplarsvsnavefecmpspashrralspa 120
 61 VQRGDVAFDAFALRSTRGHEGLAEVLEPLARSDVSNAVEFCMPSPASHRRALSPA 120
 QY 61 VQRGDVAFDAFALRSTRGHEGLAEVLEPLARSDVSNAVEFCMPSPASHRRALSPA 120
 Db 121 gytsptrvhdsvsvsfstsygdlysrarsrsralhsdthmssppvnaafpsqps 180
 121 GYTSPTRVHDSSVSFSSTSYGDLYSRARSRSRALHSDDHNSPPVNAFPSPQPS 180
 QY 121 GYTSPTRVHDSSVSFSSTSYGDLYSRARSRSRALHSDDHNSPPVNAFPSPQPS 180
 Db 181 ansstfgcsllgysssrnsfskasgptqyifheedmfvdapltisrvfdektmyrnfs 240
 181 ANSSTFGCSLLGYSSSRNSFSKASGPTQYIFHEEDMFVDAPLTISRVEDEKTMYNFS 240
 QY 181 ANSSTFGCSLLGYSSSRNSFSKASGPTQYIFHEEDMFVDAPLTISRVEDEKTMYNFS 240
 Db 241 prgmcllinnehfeqmptrngtkadkdnltlfrcmgyvickdnltgrgmlltridfak 300
 241 PRGMCLLINNEHEQMPTRNGTKADKDNLTNLFRCMGYVICKDNLTRGMLLTRIDFAK 300
 QY 241 PRGMCLLINNEHEQMPTRNGTKADKDNLTNLFRCMGYVICKDNLTRGMLLTRIDFAK 300
 Db 301 heshgsallvllshgeenvllgvdipistheilydllaanaaprlanpkivfoacrg 360
 301 HESHGSALLVLLSHGEENVLLGVDDIPISTHEIYDLLAANAAPRLANPKIVFOACRG 360
 QY 301 HESHGSALLVLLSHGEENVLLGVDDIPISTHEIYDLLAANAAPRLANPKIVFOACRG 360
 Db 361 errdngfprlvdsvdgyvafirgwdnrdgplfnflgcvrpgvqvwirkpsqadlilrya 420

Query Match 94.6%; Score 476; DB 9; Length 503;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 evl1akqylnsdngdmscgtvrekreivkavgrirgvdafdyd1rstqheglav 87
 |||||||
 Qy 28 EVL1AKQYVLSNNGDMINSCGTVREKREIVKAVGRIRGVDADYD1RSTGHEGLAV 87
 |||||||
 Db 88 leplarsvdsnavefecmpspashrrsralpsagyltsplrvhrdsvssvssftsdydiys 147
 |||||||
 Qy 88 LEPLARSDVSNVAFECMPSPASHRRSRALSPAGYTSPLRVHRDSSVSSVSSFTSYQDIYS 147
 |||||||
 Db 148 rrrrsrsralhsdrrhyspprvnafpsqpsansstfgcsajysssrnrsfskaagp 207
 |||||||
 Qy 148 RRRRSRSRALHSDRRHNSPPRVNAFPSPQSPANSSTFGCSAJYSSSRNRSFSKAGP 207
 |||||||
 Db 208 tgyifhfeedmfvdpatisrvfdektmyrnfsprgmcll1nnenhfegmptnrgtkadkd 267
 |||||||
 Qy 208 TOYIFHEEDMNFVDAPITISRVFDEKTMRYNFSPPRGKCL1NNENHFEGMPTNRGTKADKD 267
 |||||||
 Db 268 nltnlfrcmgytvlckdhl1tgrmll1trdfakshesgdasallvllshgeenvli1gvddi 327
 |||||||
 Qy 268 NLTNLFRCMGYTVICKDNLTRGMLLTRDFAKHESGDASALLVLSHGEENVLI1GVDDI 327
 |||||||
 Db 328 pistheiyd1l1naaaprlankpkivfygacgrerrdngfpyladvdvaf1rrgwdnr 387
 |||||||
 Qy 328 PISTHEIYD1L1NAANAPRLANKPKIVFYQACGRERRDNGFPYLDSDVPALFRGMWNR 387
 |||||||
 Db 368 dgplfnf1gcyrvpvgvwrkksqad1l1ryatagvsvwnsrgswf1gavcevfst 447
 |||||||
 Qy 368 DGPLFNFGCYRVPVQVWRKKPSQAD1L1RYATAGVSVWNSRGSWF1QAVCEVFST 447
 |||||||
 Db 448 hdkmdvveall1tevvkkyacgfqtsqsnllkqmpemsrllkkyf1peaarnsav 503
 |||||||
 Qy 448 HAKMDVVELL1TEVVKKYACGFQTSQSNLLKQMPEMSRLLKRYF1PEARNSAV 503
 |||||||

RESULT 6
 ID R45327 standard; Protein: 503 AA.
 AC R45327;
 DT 01-JUL-1994 (first entry)

DE Ced-3 mutant Y493.
 KW C.elegans; cell death; gene: ced-3; inhibition; human; parasite;
 KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 KW Caenorhabditis elegans.
 FT Key Location/Qualifiers
 FT misc_difference 493
 FT /label= Y493

PD MO9325694-A.
 PD 23-DEC-1993.
 PE 14-JUN-1993; U05705.
 PE 12-JUN-1992; US-897788.
 PE 20-NOV-1992; US-984182.
 PE (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PE Horvitz HR, Shahan S, Yuan J;
 PE 1993; 24-007551/01.
 PE Abstracts which affect activity of cell death genes - used to
 PE develop drugs for treating conditions characterised by cell death
 PE or proliferation
 PE Claim 99; 132pp; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC 1-beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3

CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC death such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.

SQ Sequence 503 AA:
 Query Match 92.6%; Score 466; DB 9; Length 503;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrqdrsl1ernimfshlkwdelev1iakqylnsdngdmscgtvrekreivka 60
 |||||||
 Qy 1 MMRQDRSL1ERNIMFSSHLKVDLEVLIAKQYVLSNNGDMINSCGTVREKREIVKA 60
 |||||||
 Db 61 vgrirgvdafdyd1rstqheglavleplarsvdsnavefecmpspashrrsralpsa 120
 |||||||
 Qy 61 VGRIRGVDADYD1RSTGHEGLAVLEPLARSDVSNVAFECMPSPASHRRSRALPSA 120
 |||||||
 Db 121 gysptvrhdsrvssvssftsdyqdl1ysrarsrsralhsdrrhyspprvnafpsqps 180
 |||||||
 Qy 121 GYSPTRVHRDSSVSSVSSFTSYQDIYSRARSRSRALHSDRRHNSPPRVNAFPSPQSP 180
 |||||||
 Db 181 anssffgcssl1gyssrnrsfskaagp1qyl1fheedmfvdpatisrvfdektmyrnfs 240
 |||||||
 Qy 181 ANSSFFGCSSL1GYSSRNRSFSKASGP1QY1FHEEDMNFVDAPITISRVFDEKTMRYNFS 240
 |||||||
 Db 241 prgmcll1nnenhfegmptnrgtkadkd1tnl1fcmgytvlckdhl1tgrmll1trdfak 300
 |||||||
 Qy 241 PRGMCL1NNENHFEGMPTNRGTKADKD1TNL1FRCMGYTVICKDNLTRDFAK 300
 |||||||
 Db 301 heshgdasallvllshgeenvli1gvddipistheiyd1l1naaaprlankpkivfygacg 360
 |||||||
 Qy 301 HESHGDSALLVLSHGEENVLI1GVDDIPISTHEIYD1L1NAANAPRLANKPKIVFYQACG 360
 |||||||
 Db 361 errdngfpyladvdvaf1rrgwdnr1dn1fnf1gcyrvpvgvwrkksqad1l1rya 420
 |||||||
 Qy 361 ERRDNGFPYLDSDVPALFRGMWNRDGPLFNFGCYRVPVQVWRKKPSQAD1L1RYA 420
 |||||||
 Db 421 ttgqysvsnrsgswf1gavcevfsthakmdvveall1tevvkky 466
 |||||||
 Qy 421 TTAQYVSNRSGSWF1QAVCEVFSTHAKMDVVELL1TEVVKKY 466
 |||||||

RESULT 7
 ID R45262 standard; Protein: 503 AA.
 AC R45262;
 DT 01-JUL-1994 (first entry)
 DE Ced-3.
 KW C.elegans; cell death; gene: ced-3; inhibition; human; parasite;
 KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
 KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 KW Caenorhabditis elegans.
 FT Key Location/Qualifiers
 FT peptide 1..372
 FT /note= "Claim 3, inhibitor fragment"
 FT peptide 1..149
 FT /note= "Claim 3, inhibitor fragment"
 PD MO9325694-A.
 PD 23-DEC-1993.
 PE 14-JUN-1993; U05705.
 PE 12-JUN-1992; US-897788.
 PE 20-NOV-1992; US-984182.